2017

Genetic dissection of sorghum plant architecture traits and seed number variation

Jing Zhao
Iowa State University

Genetic dissection of sorghum plant architecture traits and seed number variation

by

Jing Zhao

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Plant Breeding

Program of Study Committee:
Maria G. Salas Fernandez, Major Professor
Jianming Yu
Philip W. Becraft
Thomas Lübberstedt
Yanhai Yin

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Maria G. Salas Fernandez for giving me the opportunity to study in her group. I am grateful for her guidance and support during my Ph.D. study. I would like to extend my appreciation to the rest of committee members: Dr. Jianming Yu, Dr. Philip Becraft, Dr. Thomas Lübberstedt and Dr. Yanhai Yin for their support, guidance, and helpful suggestions.

In addition, I am thankful to the collaborators and assistants involved in my project: Dr. Ruben Peters' lab members for their work on testing enzyme activity; Dr. Kan Wang and Bin Yang for their expertise in rice transformation; Erica Unger-Wallace for her supporting in meristem dissection, Jieyun Hu, Joshua Kemp, and Riqing Li for their gracious help during the project on both lab and greenhouse activities.

Also, I would like to extend my gratefulness to all my graduate student colleagues and lab members, with whom I shared the learning process in class and field work. Special thanks to Maria Betsabe Mantilla Perez, my dear sister in the U.S., Diego Ortiz, Danielle Dykema, John Cameron, Matthew Breitzman. Thanks to our lab manager- Nicole Lindsey for her support on the project.

Finally, thanks to my family and friends back in China for their love. Thanks to my parents Qingsen Zhao and Lihua Wang for all their love and encouragement. Most significantly, I want to dedicate this to my husband Tao Zuo for being my soul partner, and for his love and support and my daughter Chelsea Zuo, my "trouble maker" giving me lots of happiness.

ABSTRACT

Sorghum [*Sorghum bicolor* (L.) Moench] is an important crop for human consumption, feed, fiber and recently biomass production. Understanding the genetic mechanisms controlling plant architecture traits will be beneficial for developing superior plant ideotypes for both grain and bioenergy production. The Sorghum Association Panel (SAP) with a remarkable diversity (>300 lines) is an excellent natural resource to dissect the genetic basis of plant architecture. We have utilized a genome-wide association study (GWAS) approach to investigate the genetic mechanisms underlying natural phenotypic variation in plant height, flowering time, panicle exsertion, panicle length, stem circumference, seed number, internode number, tiller number, leaf angle. The genotypic data used in this study included both publicly available GBS data and gene-specific markers developed by SEQUENOME for the gibberellin (GA) biosynthesis and signaling pathways. A total of 101 SNP representative regions were associated with at least one of the nine plant architecture traits investigated and two of them corresponded to GA candidate genes affecting plant height and seed number, respectively. The candidate gene *SbKS3*, associated with variation in seed number per panicle had two alternative haplotypes that corresponded to high and low seed number lines. The second project confirmed the effect of *SbKS3* on seed number and, preliminary evidence suggests that this effect is generated by a modification in panicle branch length.

CHAPTER 1. GENERAL INTRODUCTION

Sorghum Bicolor

Sorghum [*Sorghum bicolor* (L.) Moench] is a drought tolerant C_4 crop. It is widely adapted to variable environmental conditions, and ranks fifth among the most important cereal species worldwide, after rice, wheat, maize and barley. Sorghum is a versatile crop used for food, feed, fiber, and energy production (Rooney 2004; Salas Fernandez et al. 2009).

Sorghum is primarily a self-pollinating species, but it is also highly amenable to commercial hybrid seed production by the utilization of cytoplasmic genetic male sterility (CMS). Several CMS systems (*e.g*., the A_1 system, Stephens and Holland 1954; A_3 system, Tang et al. 1996) have been discovered in sorghum, and applied to exploit heterosis in hybrid production by crossing a male sterile line with a male fertile line. Sorghum is, therefore, an annual crop that can be grown as either a self-pollinated variety or a hybrid. Heterosis refers to the phenomenon in which progenies perform better than their parents in terms of growth rate, fertility, biomass, and grain yield and has been an important driving force in plant evolution (Birchler et al. 2003). Even though, the genetic determinants of heterosis are still unclear, Li et al. (2015) demonstrated a pseudo-overdominance mechanism determining heterosis in sorghum plant height, *i.e.*, two closely linked loci with dominant alleles in repulsion resulted in a perceived overdominance effect.

Sorghum domestication begun approx. 3,000 to 6,000 years ago in what is now Ethiopia and Sudan, from where it spread and diversified throughout the world (Kimber, 2000). In the United States, sorghum was first introduced around 1874-1908. Narrow founder varieties were widely planted at that time, and spontaneous mutations for early maturity and short stature were selected in the breeding program (Klein et al. 2008). By the 1960s, sorghum breeders recognized

the problem of a strong genetic bottleneck in the US breeding programs (Smith et al. 2010). Even though there were diverse accessions outside of the U.S., most of these could not be utilized in temperate climates due to their photoperiod sensitivity and extreme height. Therefore, the Sorghum Conversion Program (SCP), a strategy to introduce tropical germplasm into non-tropical environments, was initiated at that time. In general, the program was based on a breeding scheme that included an initial cross between tropical accessions and a temperate elite line (BTx406) in Puerto Rico, followed by phenotypic progeny selection based on reduced height and day-neutral flowering response in Texas. The selected progenies were then backcrossed to the tropical progenitor in Puerto Rico, up to five times. The final goal of the SCP was to generate a collection of sorghum converted (SC) lines with up to 99% of the genes from tropical progenitors and the introgressed height and maturity genes from BTx406 (Stephens et al. 1967). To investigate the consequences of the SCP, Thurber et al. (2013) compared genotypic information of sorghum converted lines and tropical progenitors. As expected, sorghum converted lines had high genotypic similarity to their exotic progenitors, but dramatic genotypic differences in plant height (*Dw*$_{1\text{-}4}$) and flowering time loci (*Ma*$_{1\text{-}6}$). This resource has been very useful to enrich sorghum breeding programs in the U.S., and offers a great opportunity to dissect the genetic bases of agricultural important traits.

Sorghum bicolor is a diploid species with a base chromosome number of n=10. The reference genome was obtained from BTx623 with a size of approx. 736Mb, which is larger than rice but substantially smaller than maize and other grass species (Paterson et al. 2009). According to Paterson et al. (2009), the sorghum enrichment in retrotransposons is the major explanation of a larger genome relative to rice genomes of other grass species, such as maize or sugarcane, are larger and more complex mainly due to gene duplications. For instance, maize

diverged from sorghum ~50 million years ago (Akhani et al., 2012) and has undergone a whole genome duplication that led to a larger genome size (2500 Mb) (Osborne and Beerling 2006). Sugarcane, an important feedstock for biofuel production worldwide, diverged from sorghum only 5 million years ago (Sage and Monson 1999) and has subsequently experienced at least two duplications (Sage 2004) that explain the 10 Gb size of its genome. From the physiological perspective, sorghum is the representative of tropical grasses that utilize the C_4 photosynthesis pathway, while rice is the preferred model species for C_3 photosynthesis. Altogether, the small genome size and limited gene duplication rate makes sorghum an attractive model species for functional genomics of C_4 plants.

Importance of Plant Architecture

Plant architecture is the three-dimensional organization of the plant body, related to the disposition of vegetative organs that capture light, the synchrony of inflorescence, seed development, and thus, grain production. The above ground architecture, including branching pattern, size, shape, and position of leaf and flower organs, has long been the basis of taxonomic classifications and is an important determinant of yield potential (Reinhardt and Kuhlemeier 2002).

Notably, the “Green Revolution” that significantly increased wheat (Peng et al. 1999) and rice (Ikeda et al. 2001) productivity, was based on the modification of plant height, a basic component of plant architecture. Leaf angle, another important plant architecture trait, has also been extensively exploited to increase biomass and grain yield. Erect leaves not only improve the efficiency of light capture for photosynthesis and influence nitrogen reservoirs for grain filling, but also enable denser plantings with a higher leaf area index per unit of land (Sinclair and Sheehy 1999). Therefore, reducing leaf angle has generated an increase in biomass and grain

yield, as reported in rice (Shimada et al. 2006), maize (Stewart et al. 2003), wheat, oat and barley (Tanner et al. 1966). Sorghum has abundant morphological diversity from natural variation and artificial selections, making it an outstanding model to dissect the genetic control of plant architecture.

Importance of Seed Number as a Yield Component Trait

Seed number or number of grains per panicle, is not only a plant architecture trait but also an important yield component. Grain yield in cereal crops is determined by four primary components: number of plants per unit of land, number of panicles per plant, seed number per inflorescence and seed size (weight per seed) (Egli 2017). "Number of plants per unit of land" refers to plant density and could be improved by modifying plant architecture traits. Even though the latter three components determine yield on a per plant basis, the "number of panicles per plant" does not contribute much to final yields because of the additional resources needed to produce secondary stems (Egli 2017). Therefore, seed number per inflorescence and seed size are the two components contributing the most to total grain yield, as indicated by Heinrich et al. (1983). Several studies across species concluded that seed number is the most important determinant of yield (Saeed et al. 1986; Gerik et al. 2003; Peltonen-Sainio et al. 2007). This discovery was validated by genetic studies, in which the modification of genes controlling seed number resulted in enhanced grain yield (Ashikari et al. 2005; Wu et al. 2016).

However, there is evidence of a tradeoff between seed number per panicle and seed size (Jakobsson and Eriksson, 2000; Paul-Victor and Turnbull 2009; Sadras 2007), a phenomenon that causes an overall neutral effect on grain yield, since an increase in seed number is usually accompanied by a decrease in seed size. For example, the application of the auxin transport inhibitor 2,3,5-triiodobenzoic acid (TIBA) on soybean caused an increase in seed number with a

concomitant decrease in seed size and thus, no effect on final yield (Tanner and Ahmed 1974). Burow et al. (2014) reported similar results in sorghum in which seed number was tripled in an Ethyl methane sulfonate (EMS)-induced mutant, but yield was reduced due to a decrease in seed size. These observations can be, at least partly, explained by the "Source-Sink Limitations", determined by the capacity of the plant community to generate assimilates through photosynthesis (the source) and the capacity of the seeds to use these assimilates (the sink). According to Murata (1969), seed number per inflorescence is determined by the carbon fixation capacity during a critical period that includes the development of reproductive structures, pollination, fertilization and initial stages of seed growth. A variation in the photosynthetic capacity during this critical period usually results in a corresponding change in seed number (Fischer and Laing, 1976; Nico et al. 2016). The "Golden Rule" of crop physiology, in which maximum yield requires maximum solar radiation interception early during reproductive growth, is based on the source control of seed number and yield (Egli, 2017). On the other hand, even if the carbon fixation capacity is not altered, seed number per inflorescence can be improved by reducing the plant biomass. This approach was clearly successful during the Green Revolution; *e.g.*, IR8, the semi-dwarf variety released at that time produces more grains per rice plant when grown with irrigation and fertilizer treatments. Therefore, it is clear that the ultimate goal of increasing seed number can be obtained through alternative biological mechanisms.

The Role of Gibberellins on Plant Architecture and Seed Number Determination

Plant hormones, including auxins, gibberellins (GAs), cytokinins, ethylene, abscisic acid and brassinosteroids (BR) play an essential role in plant growth, development and architecture (Taiz and Ziger, 2010). Gibberellins were the second group of hormones discovered in rice after

auxins, and are typically involved in germination, cell division and elongation. Additionally, gibberellins have been extensively reported to play an important role in plant architecture traits, *i.e*., leaf angle (Shimada et al. 2006), tillering (Lo et al. 2008), stem diameter (Leite et al. 2003), panicle exsertion (Gao et al. 2016), and internode number (Jupe et al. 1988). In addition to plant architecture traits, GAs have an effect on seed development by stimulating pollen development and pollen tube growth, promoting fruit set, and reducing seed abortion (Swain and Singh 2005).

The GA biosynthesis and signaling pathways have been broadly studied in model plants such as Arabidopsis and rice. The synthesis of GAs starts in embryos during seed imbibition (Kaneko et al. 2003), and then continues in many parts of the plants such as young seedlings, shoot apices, and developing seeds. GAs are synthesized by the terpenoid pathway that can be divided into three main stages. In the first stage, geranylgeranyl diphosphate (GGPP) is converted into a tetracyclic compound, *ent*-kaurene in cytoplast. In stage II, which occurs on the plastid envelope and in the endoplasmic reticulum, the reaction is followed by the conversion of *ent*-kaurene to GA12. During stage III, GA12 is finally converted into other C20-GA in the cytosol, and then into C19-GA including active GAs. Endogenous bioactive GAs regulate their own synthesis by enhancing or inhibiting the transcription of GA biosynthesis or deactivation genes (Sakamoto et al. 2004; Yamaguchi 2008). The signaling pathway begins with the binding of GA to the receptor GID1, which then stimulates the formation of GA-GID1-DELLA complexes. F-box proteins, components of the SKP1-CULLIN-F-BOX (SCF) E3 ubiquitin-ligase complex, catalyze the attachment of polyubiquitin to DELLA proteins that in turn, trigger the degradation of the GA-GID1-DELLA complex by 26 proteosomes. DELLA proteins function as stimulating factors by a dual mechanism: (i) in the absence of the SCF complex, the GA-GID1

complex binds to DELLAs inhibiting their function, and (ii) in the presence of the SCF complex, DELLAs are released from the GA-GID1-DELLA complex (Davière and Achard 2013).

GAs do not act alone; other hormones such as brassinosteroids and cytokinins, interact with them in diverse ways contributing to their powerful roles in plant architecture and seed development. E.g., the rice *SPINDLY* (SPY) gene functions as a negative regulator of gibberellin signaling, and transgenic rice plants transformed with an OsSPY RNAi construct showed a larger leaf angle by modulating brassinosteriod synthesis (Shimada et al. 2006). *Grain Number per Panicle1* (*GNP1*) encodes GA20ox1, a gibberellin biosynthetic enzyme, and controls grain number and yield through a mechanism that involves a KNOX-mediated cytokinin and gibberellin crosstalk in the rice meristem (Wu et al. 2016).

Genome-Wide Association Studies

Prior to the introduction of association studies, linkage mapping using biparental populations was the primary method of quantitate genetic analysis. This approach proved highly useful, but the results are usually hard to reproduce for common and complex traits. Genetic association studies, including genome-wide association studies (GWAS) and candidate-gene approaches, emerged as an alternative and useful complement to linkage mapping. Briefly, this method evaluates whether certain alleles within a population are correlated with specific phenotypes more or less often than expected, exploiting ancestral recombination in natural populations (Buckler and Thornsberry 2002). GWAS has become a widely-used method due to its multiple advantages. GWAS-based discoveries can be easier to reproduce than linkage mapping results because GWAS investigate common alleles while linkage mapping often targets rare alleles. GWAS can better capture the diversity of a species by testing large collections of individuals with distinct genetic backgrounds. The mapping resolution can be greatly increased

due to a longer recombination history captured on diversity panels than on designed bi-parental populations, typically used in linkage mapping studies (Shehzad et al. 2009; Myles et al. 2009; Huang and Han 2014). Finally, GWAS can help identify novel genomic regions associated with a phenotype in the absence of previous knowledge about the trait, in comparison with candidate gene approaches. The major concern in a GWAS is the number of false positive associations that could be obtained as a consequence of population structure and within-group familial relatedness (Yu and Buckler 2006). Numerous statistical methods have been designed to minimize these spurious associations. Yu et al. (2006) proposed a unified mixed model approach, in which population structure (Q) and a kinship matrix (K) are estimated using random markers, and then fit into a mixed model to test for marker-trait associations.

Genome wide association studies have been successfully used for gene/genomic region discovery in multiple species: *Arabidopsis thaliana* (L.) Heynh. (Atwell et al. 2010; Sandra et al. 2016), rice (*Oryza sativa L.;* Huang et al. 2010; Yano et al. 2016), maize (*Zea mays L.*) (Li et al. 2013; Tian et al. 2011), oat (*Avena sativa L.*) (Giorgio et al. 2016; Newell et al. 2012), and barley (*Hordeum vulgare L.)* (Matthies et al. 2014; Pasam et al. 2012). In sorghum, a few GWAS have been performed to investigate different plant characteristics using a sorghum diversity panel (Brown et al. 2008; Shehzad et al. 2009) and a mini core panel (Upadhyaya et al. 2012a; Upadhyaya et al. 2012b) with limited genome coverage. A GWAS using high-throughput genotyping data and a large sorghum association panel was reported by Morris et al. (2013) for plant height and inflorescence branch length. A similar approach was used to discover the genetic loci underlying phenotypic variation in plant height and inflorescence architecture (Zhang et al. 2015), plant height and heterosis (Li et al. 2015), and yield components (Boyles et al. 2016). Loci consistently identified by both linkage mapping approaches and GWAS have

been reported for plant height on sorghum chromosomes 6 and 9 (Brown et al. 2008; Morris et al. 2013; Zhang et al. 2015; Li et al. 2015), and other traits (Boyles et al. 2016; Zhang et al. 2015).

Association studies have been recently demonstrated as an effective method for gene discovery since genes controlling two complex traits, seed size and drought tolerance, were successfully cloned after an initial discovery by GWAS. *OsSPL13* was first identified using a GWAS and its function later confirmed controlling seed size in rice (Si et al. 2016). Similarly, *ZmVPP1* that confers drought tolerance in maize seedlings was cloned and its function confirmed by a transformation study after an initial GWAS (Wang et al. 2016). These successful cases suggest that gene cloning could be the final outcome of a gene discovery process that starts with quantitative genetic approaches that utilize large and diverse populations.

Goals and Objectives

The main goal of this project was to elucidate the genetic control of plant architecture traits and seed number variation, a critical yield component in sorghum that will benefit future breeding efforts to increase productivity.

Aim 1. To conduct a genome-wide discovery of chromosomal regions and candidate genes controlling plant height, flowering time, panicle exsertion, panicle length, leaf angle, stem circumference, tiller number, internode number, and seed number per inflorescence, and investigate the association between allelic variation in GA genes and our traits of interest.

***Aim* 2.** To validate and characterize a hormonal gene identified in Aim 1 that controls seed number per inflorescence.

References

Ashikari M, Sakakibara H, Lin S, et al (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–5. doi: 10.1126/science.1113373Atwell S, Huang YS, Vilhjálmsson BJ, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465:627–631. doi: 10.1038/nature08800

Birchler JA, Auger DL, Riddle NC (2003) In search of the molecular basis of heterosis. Plant Cell 15:2236–2239. doi: 10.1105/tpc.151030

Bortesi L, Fischer R (2015) The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol Adv 33:41–52. doi: 10.1016/j.biotechadv.2014.12.006

Boyles RE, Cooper EA, Myers MT, et al (2016) Genome-Wide Association Studies of Grain Yield Components in Diverse Sorghum Germplasm. Plant Genome 0:0. doi: 10.3835/plantgenome2015.09.0091

Sandra E, Branham S, Wright SJ, Reba A, et al (2016) Genome-Wide Association Study in Arabidopsis thaliana of Natural Variation in Seed Oil Melting Point: A Widespread Adaptive Trait in Plants. J Hered.107(3): 257–265. doi: 10.1093/jhered/esw008

Bhalla PL (2006) Genetic engineering of wheat – current challenges and opportunities. TRENDS in Biotechnology. 24:305-311. doi: 10.1016/j.tibtech.2006.04.008

Brown PJ, Klein PE, Bortiri E, et al (2006) Inheritance of inflorescence architecture in sorghum. Theor Appl Genet 113:931–942. doi: 10.1007/s00122-006-0352-9

Brown PJ, Rooney WL, Franks C, Kresovich S (2008) Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes Genetics 180:629–637 doi: 101534/genetics108092239

Buckler IV ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111. doi: 10.1016/S1369-5266(02)00238-8

Burow G, Xin Z, Hayes C, Burke J (2014) Characterization of a multiseeded (msd1) mutant of sorghum for increasing Grain Yield. Crop Sci 54:2030–2037. doi: 10.2135/cropsci2013.08.0566

Cyr A, Wilderman PR, Determan M, Peters RJ (2007) A modular approach for facile biosynthesis of labdane-related diterpenes. J Am Chem Soc 129:6684–6685. doi: 10.1021/ja071158n

Davière J-M, Achard P (2013) Gibberellin signaling in plants. Development 140:1147–51. doi: 10.1242/dev.087650

Davuluri, GR, et al. (2005) Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and favonoid content in tomatoes. Nat. Biotechnol. 23, 890–895

De Block M, et al. (1997) The development of a nuclear male sterility system in wheat. Expression of the barnase gene under the control of tapetum-specific promoters. Theor. Appl. Genet. 95, 125–131 9

Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science (80-) 346:1258096–1258096. doi: 10.1126/science.1258096

Egli DB (2017) Seed Biology and Yield of Grain Crops, 2nd Edition. University of Kentukey, USA. ISBN 13:9781780647708

Fischer RA, and DR Laing (1976) Yield potential in a dwarf spring wheat and response to crop thinningSource. The Journal of Agricultural Science 87:113-122

Gao S, Fang J, Xu F, et al (2016) Rice HOX12 Regulates Panicle Exsertion by Directly Modulating the Expression of ELONGATED UPPERMOST INTERNODE1. Plant Cell 28:tpc.01021.2015. doi: 10.1105/tpc.15.01021

Gerik, T., B. W. Bean, and R. Vanderlip. 2003. Sorghum Growth and Development. Agric. Communications, The Texas A&M Univ. Syst., College Station, TX.

Heinrich GM, Francis C a., Eastin JD (1983) Stability of Grain Sorghum Yield Components Across Diverse Environments1. Crop Sci 23:209. doi: 10.2135/cropsci1983.0011183X002300020004x

Higgins RH, Thurber CS, Assaranurak I, Brown PJ (2014) Multiparental mapping of plant height and flowering time QTL in partially isogenic sorghum families. G3 (Bethesda) 4:1593–602. doi: 10.1534/g3.114.013318

Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. Annu Rev Plant Biol 65:531–51. doi: 10.1146/annurev-arplant-050213-035715

Huang X, Wei X, Sang T, et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967. doi: 10.1038/ng.695

Ikeda A, Ueguchi-Tanaka M, Sonoda Y, et al (2001) slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. Plant Cell 13:999–1010. doi: 10.1105/tpc.13.5.999

Jakobsson A, Eriksson O (2000) A comparative study of seed number, seed size, seedling size and recruitment in grassland plants. OIKOS 88: 494–502. DOI: 10.1034/j.1600-0706.2000.880304.

Jupe SC, Causton DR, Scott IM (1988) Cellular basis of the effects of gibberellin and the pro gene on stem growth in tomato. Planta 174:106–111. doi: 10.1007/BF00394881

Kajala K, Covshoff S, Karki S, et al (2011) Strategies for engineering a two-celled C4 photosynthetic pathway into rice. J Exp Bot 62:3001–3010. doi: 10.1093/jxb/err022

Kaneko M, Itoh H, Inukai Y, et al (2003) Where do gibberellin biosynthesis and gibberellin signaling occur in rice plants? Plant J 35:104–115. doi: 10.1046/j.1365-313X.2003.01780.x

Klein RP, Mullet JE, Jordan DR, Miller FR, Rooney WL, et al (2008) The eff ect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Sci. 48(S1):S12–S26

Kramer MG, Redenbaugh K (1994) Commercialization of a tomato with an antisense polygalacturonase gene: The FLAVR SAVR™ tomato story. Euphytica. **79** (3): 293–297. *ISSN 0014-2336*. *doi*:*10.1007/BF00022530*.

Leite VM, Rosolem CA, Rodrigues JD (2003) Gibberellin and cytokinin effects on soybean growth. Sci Agric 60:537–541. doi: 10.1590/S0103-90162003000300019

Li Hui, Peng ZY, Yang XH, Wang WD (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nature Genetics 45: 43-50. doi:10.1038/ng.2484

Li X, Li X, Fridman E, et al (2015) Dissecting repulsion linkage in the dwarfing gene *Dw3* region for sorghum plant height provides insights into heterosis. Proc Natl Acad Sci 112:11823–11828. doi: 10.1073/pnas.1509229112

Lo S, Yang S, Chen K, et al (2008) A Novel Class of Gibberellin 2-Oxidases Control Semidwarfism , Tillering , and Root Development in Rice. 20:2603–2618. doi: 10.1105/tpc.108.060913

Lowe K, Wu E, Wang N, et al (2016) Morphogenic Regulators Baby boom and Wuschel Improve Monocot Transformation. Plant Cell 28:tpc.00124.2016. doi: 10.1105/tpc.16.00124

Matthies IE, Malosetti M, Röder MS, van EF (2014) Genome-Wide Association Mapping for Kernel and Malting Quality Traits Using Historical European Barley Records. PLOS ONE 9(11): e110046. https://doi.org/10.1371/journal.pone.0110046

Miura K, Ikeda M, Matsubara A, et al (2010) OsSPL14 promotes panicle branching and higher grain productivity in rice. Nat Genet 42:545–549. doi: 10.1038/ng.592

Morris GP, Ramu P, Deshpande SP, et al (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci 110:453–8. doi: 10.1073/pnas.1215985110

Morrone D, Lowry L, Determan MK, et al (2010) Increasing diterpene yield with a modular metabolic engineering system in E. coli: Comparison of MEV and MEP isoprenoid precursor pathway engineering. Appl Microbiol Biotechnol 85:1893–1906. doi: 10.1007/s00253-009-2219-x

Multani DS, Briggs SP, Chamberlin MA, et al (2003) Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. Science (80-) 302:81–84. doi: 10.1126/science.1086072

Murata Y (1969) Physiological responses to nitrogen in plants. Agron Publ 197.

Murphy RL, Klein RR, Morishige DT, et al (2011) Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. Proc Natl Acad Sci U S A 108:16469–16474. doi: 10.1073/pnas.1106212108

Myles S, Peiffer J, Brown PJ, et al (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202. doi: 10.1105/tpc.109.068437

Nagaraja Reddy R, Madhusudhana R, Murali Mohan S, et al (2013) Mapping QTL for grain yield and other agronomic traits in post-rainy sorghum [Sorghum bicolor (L.) Moench]. Theor Appl Genet 126:1921–1939. doi: 10.1007/s00122-013-2107-8

Newell MA, Asoro FG, Scott MP, et al (2012) Genome-wide association study for oat (Avena sativa L.) beta-glucan concentration using germplasm of worldwide origin. Theor Appl Genet 125:1687–1696. doi: 10.1007/s00122-012-1945-0

Nico M, Mantese AI, Miralles DJ, Kantolic AG (2016) RESEARCH PAPER Soybean fruit development and set at the node level under combined photoperiod and radiation conditions. Journal of Experimental Botany 67:365–377. doi:10.1093/jxb/erv475

Paul-Victor C, Turnbull L (2009) The effect of growth conditions on the seed size/number trade-off. PLoS One 4:e6917

Pasam RK, Sharma R, Malosetti M, et al (2012) Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol 12:16. doi: 10.1186/1471-2229-12-16

Paterson AH, Bowers JE, Bruggmann R, et al (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556. doi: 10.1038/nature07723

Peltonen-Sainio P, Kangas A, Salo Y, Jauhiainen L (2007) Grain number dominates grain weight in temperate cereal yield determination: Evidence based on 30 years of multi-location trials. F Crop Res 100:179–188. doi: 10.1016/j.fcr.2006.07.002

Rami J-F, Dufour P, Trouche G, et al (1998) Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (Sorghum bicolor L. Moench). TAG Theor Appl Genet 97:605–616. doi: 10.1007/s001220050936

Reinhardt D, Kuhlemeier C (2002) Plant architecture. EMBO Rep 3:846–851. doi: 10.1093/embo-reports/kvf177

Rooney WL (2004) Sorghum Improvement - Integrating Traditional and New Technology to Produce Improved Genotypes. Adv Agron 83:37–109. doi: Doi 10.1016/S0065-2113(04)83002-5

Sadras VO (2007) Evolutionary aspects of the trade-off between seed size and number in crops. F Crop Res 100:125–138. doi: 10.1016/j.fcr.2006.07.004

Saeed M, Francis CA, Clegg MD (1986) Yield Component Analysis in Grain Sorghum. Crop Sci 26:346–351. doi: 10.2135/cropsci1986.0011183X002600020028x

Sakamoto T, Miura K, Itoh H, et al (2004) An Overview of Gibberellin Metabolism Enzyme Genes and Their Related Mutants in Rice 1 [w]. 134:1642–1653. doi: 10.1104/pp.103.033696.two

Sakamoto T, Morinaka Y, Ohnishi T, et al (2006) Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice. Nat Biotechnol 24:105–109. doi: 10.1038/nbt1173

Salas Fernandez MG, Becraft PW, Yin Y, Lübberstedt T (2009) From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci 14:454–461. doi: 10.1016/j.tplants.2009.06.005

Shehzad T, Iwata H, Okuno K (2009) Genome-wide association mapping of quantitative traits in sorghum (Sorghum bicolor (L) Moench) by using multiple models. Breed Sci 59:217–227 doi: 101270/jsbbs59217

Shehzad T, Iwata H, Okuno K (2009) Genome-wide association mapping of quantitative traits in sorghum (Sorghum bicolor (L) Moench) by using multiple models. Breed Sci 59:217–227. doi: 10.1270/jsbbs.59.217

Shimada A, Ueguchi-Tanaka M, Sakamoto T, et al (2006) The rice SPINDLY gene functions as a negative regulator of gibberellin signaling by controlling the suppressive function of the DELLA protein, SLR1, and modulating brassinosteroid synthesis. Plant J 48:390–402. doi: 10.1111/j.1365-313X.2006.02875.x

Si L, Chen J, Huang X, et al (2016) OsSPL13 controls grain size in cultivated rice. Nat Genet 48:447–456. doi: 10.1038/ng.3518

Sreenivasulu N, Schnurbusch T (2012) A genetic playground for enhancing grain number in cereals. Trends Plant Sci 17:91–101. doi: 10.1016/j.tplants.2011.11.003

Stephens JC, Holland RF (1954) Cytoplasmic Male-Sterility For Hybrid Sorghum Seed Production. Agron J 46:20–23. doi: 10.2134/agronj1954.00021962004600010006x

Stephens JC, Miller FR, Rosenow DT (1967) Conversion of Alien Sorghums to Early Combine Genotypes. Crop Sci 7:396. doi: 10.2135/cropsci1967.0011183X000700040036x

Stewart DW, Costa C, Dwyer LM, et al (2003) Canopy Structure, Light Interception, and Photosynthesis in Maize. Agron J 95:1465–1474. doi: 10.2134/agronj2003.1465

Swain SM, Singh DP (2005) Tall tales from sly dwarves: Novel functions of gibberellins in plant development. Trends Plant Sci 10:123–129. doi: 10.1016/j.tplants.2005.01.007

Tanner JW, Ahmed S (1974) Growth analysis of soybeans treated with TIBA. Crop Sci 14:371–374.

Tanner JW, Gardener CJ, N.C S, Reinbergs.E (1966) Some observation on upright . leaf. type small grains. CanJ Plant Sci 46:1966.

Thurber CS, Ma JM, Higgins RH, Brown PJ (2013) Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol 14:R68. doi: 10.1186/gb-2013-14-6-r68

Giorgio Tumino G, Voorrips RE, Rizza F, Badeck FW (2016) Population structure and genome-wide association analysis for frost tolerance in oat using continuous SNP array signal intensity ratios. Theor Appl Genet 129: 1711. doi:10.1007/s00122-016-2734-y

Van Oosterom EJ, Hammer GL (2008) Determination of grain number in sorghum. F Crop Res 108:259–268. doi: 10.1016/j.fcr.2008.06.001

Wang X, Wang H, Liu S, et al (2016) Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. Nat Genet 48:1–12. doi: 10.1038/ng.3636

Wei X, Xu J, Guo H, et al (2010) DTH8 Suppresses Flowering in Rice, Influencing Plant Height and Yield Potential Simultaneously. Plant Physiol 153:1747–1758. doi: 10.1104/pp.110.156943

Witt Hmon KP, Shehzad T, Okuno K (2014) QTLs underlying inflorescence architecture in sorghum (Sorghum bicolor (L.) Moench) as detected by association analysis. Genet Resour Crop Evol 61:1545–1564. doi: 10.1007/s10722-014-0129-y

Wu Y, Wang Y, Mi XF, et al (2016) The QTL GNP1 Encodes GA20ox1, Which Increases Grain Number and Yield by Increasing Cytokinin Activity in Rice Panicle Meristems. PLoS Genet 12:1–21. doi: 10.1371/journal.pgen.1006386

Xue W, Xing Y, Weng X, et al (2008) Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat Genet 40:761–767. doi: 10.1038/ng.143

Yamaguchi M, Fujimoto H, Hirano K, et al (2016) Sorghum Dw1, an agronomically important gene for lodging resistance, encodes a novel protein involved in cell proliferation. Sci Rep 6:28366. doi: 10.1038/srep28366

Yamaguchi S (2008) Gibberellin Metabolism and its Regulation. Annu Rev Plant Biol 59:225–251. doi: 10.1146/annurev.arplant.59.032607.092804

Yano K, Yamamoto E, Aya K, Takeuchi H, et al (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice**.** Nature Genetics 48:927-934. doi:10.1038Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155–160. doi: 10.1016/j.copbio.2006.02.003

Zhang D, Li JP, Rosana O. Compton, JR, Goff VH, et al (2015) Comparative Genetics of Seed Size Traits in Divergent Cereal Lineages Represented by Sorghum (Panicoidae) and Rice (Oryzoidae). G3: GENES, GENOMES, GENETICS. 5:1117-1128. *doi.org/10.1534/g3.115.017590*

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler E (2010) Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42:355-360.

Zhang, Y. et al. (2005) Agrobacterium–mediated transformation and generation of male sterile lines of Australian canola. Aust. J. Agric. Res. 56, 353–361

Zhao J, Mantilla Perez MB, Hu J, Salas Fernandez MG (2016) Genome-Wide Association Study for Nine Plant Architecture Traits in Sorghum. Plant Genome 9:0. doi: 10.3835/plantgenome2015.06.0044

CHAPTER 2. GENOME WIDE ASSOCIATION STUDY FOR NINE PLANT ARCHITECTURE TRAITS IN *SORGHUM BICOLOR*

Modified from a paper published in The Plant Genome 9(2):1-14 doi:

10.3835/plantgenome2015.06.0044

Jing Zhao[1], Maria B. Mantilla Perez[1], Jieyun Hu[1], Maria G. Salas Fernandez[1]*

[1]Department of Agronomy, Iowa State University, Ames, IA 50010.

Abstract

Sorghum [*Sorghum bicolor* (L) Moench], an important gain and forage crop, is receiving significant attention as a lignocelullosic feedstock because of its water-use efficiency and high biomass yield potential. Because of the advancement of genotyping and sequencing technologies, genome-wide association study (GWAS) has become a routinely used method to investigate the genetic mechanisms underlying natural phenotypic variation. In this study, we performed a GWAS for nine grain and biomass-related plant architecture traits to determine their overall genetic architecture and the specific role of natural allelic variations in gibberellin (GA) biosynthesis and signaling genes with these phenotypes. A total of 101 Single-nucleotide polymorphism (SNP) representative regions were associated with at least one of the nine traits and two of the significant markers correspond to GA candidate genes, *GA2ox5 (Sb09g028360)* and *KS3 (Sb06g028210),* affecting plant height and seed number respectively. The resolution of a previously reported quantitative trait loci (QTL) for leaf angle on chromosome 7 was increased to a 1.67 Mb region containing seven candidate genes with good prospects for further investigation. This study provides new knowledge of the association of GA genes with plant architecture traits and the genomic region controlling variation in leaf angle, stem circumference, internode number, tiller number, seed number, panicle exsertion and panicle length. The GA

gene affecting seed number variation (*KS3*, *Sb06g028210*) and the genomic region on chromosome 7 associated with variation in leaf angle are also important outcomes of this study and represent the foundation for future validation studies needed to apply this knowledge in breeding programs.

Introduction

The increasing interest in biomass production for biofuel use is resulting in a paradigm shift in breeding for plant architecture parameters. The genetic manipulation of these traits can positively affect biomass production (Yuan et al. 2008) as suggested by the high correlations between biomass yield and plant height (Lübberstedt et al. 1997;Salas Fernandez et al. 2009) or leaf angle (Morinaka et al. 2006). Sorghum, the fifth most widely grown cereal crop in the world, is receiving significant attention as one of the most productive annual species for bioenergy production (Rooney et al. 2007) in addition to its well-known value as a grain and forage crop. Therefore, understanding the genetic control of plant architecture traits and applying that knowledge in sorghum breeding programs might be instrumental to develop improved germplasm for the incipient lignocellulosic feedstock market as well as contribute to increase yield in grain and forage sorghum breeding programs.

Several linkage mapping studies have been conducted in sorghum to dissect the genetic mechanisms controlling plant architecture. Traits such as plant height, flowering time, and panicle length have been characterized in different segregating populations (Hart et al. 2001;Nagaraja Reddy et al. 2013;Srinivas et al. 2009; Zuo et al. 2012). Other traits such as panicle exsertion (Feltus et al. 2006; Klein et al. 2001), tiller number (Alam et al. 2014; Feltus et al. 2006; Hart et al. 2001; Murray et al. 2008; Paterson et al. 1995; Shiringani et al. 2010), and internode number (Nagaraja Reddy et al. 2013; Srinivas et al. 2009; Zou et al. 2012) have also

been investigated by several groups. However, limited information is available for leaf angle (Hart et al. 2001), stem circumference (Zou et al. 2012) and seed number (Nagaraja Reddy et al. 2013). Most QTL identified in these studies were specific to a single population, a finding consistent with the nature of biparental populations but in some cases, the comparative analysis of multiple independent studies allowed for the identification of a QTL consistent across populations. Panicle length is an example, in which a QTL was identified by four groups in the region 58,285,987-61,171,968 bp on chromosome 7 (Hart et al. 2001; Brown et al. 2006; Nagaraja Reddy et al. 2013; Srinivas et al. 2009).

Association mapping, also known as Linkage Disequilibrium (LD) mapping, is a powerful tool to detect chromosomal regions controlling quantitative traits that depends on LD structure across the genome (Flint-Garcia et al. 2003). Although LD mapping could generate false positive associations between phenotype and genotype, Zhang et al. (2010) demonstrated that controlling for population structure and familial relatedness greatly reduced the number of spurious associations. The advantages of LD mapping, such as the short time span needed for population development, its broad application, and the large statistical power when used with high-throughput genotyping data, have been determinants for its frequent use in gene or marker discovery studies (Myles et al. 2009; Huang and Han 2014) in multiple species: *Arabidopsis thaliana* (L.) Heynh.(Atwell et al. 2010), rice (*Oryza sativa* L.) (Huang et al. 2010), maize (*Zea mays* L.) (Tian et al. 2011), oat (*Avena sativa* L.) (Newell et al. 2012), and barley (*Hordeum vulgare* L.) (Pasam et al. 2012). In sorghum, a few association studies have been performed to investigate specific plant characteristics using a diversity panel (Brown et al. 2008; Shehzad et al. 2009; Murray et al. 2009; Mantilla Perez et al. 2014) and a mini core panel (Upadhyaya et al. 2012a;Upadhyaya et al. 2012b) with limited genome coverage. Additionally, several GWAS

have been recently conducted using high-throughput genotyping data to discover the genetic control of grain polyphenol concentration (Rhodes et al. 2014), flavonoid pigmentation traits (Morris et al. 2013b), aluminum tolerance and grain yield in P-limited envrionments (Leiser et al. 2014), resistance to stalk rot diseases (Adeyanju et al. 2015), seed size (Zhang et al. 2015b), and plant height and inflorescence trait components (Morris et al. 2013; Zhang et al. 2015a).

In spite of the wide genome coverage obtained with methods such as genotyping-by-sequencing (GBS) (Elshire et al. 2011), genotypic data sets could be enriched for regions not covered as a result of technical limitations. The abundant information about genes controlling plant architecture traits in model species can be leveraged and applied to gene discovery studies in crop species. E.g. a dwarf gene encoding a DELLA protein has been identified as part of the GA signaling pathway and has been cloned in *Arabidopsis* (Peng et al. 1997), rice (Ikeda et al. 2001), maize (Harberd and Freeling 1989), barley (Chandler et al. 2002) and wheat (Peng et al. 1999). Therefore, this highly conserved gene is a good candidate to investigate its potential effect on plant height in sorghum. As recently reported by Ordonio et al. (2014), sorghum mutants in four GA biosynthesis genes have pleotropic effects on height and stalk erectness, and thus, the association of natural allelic variation in those four genes with plant architecture should be investigated. Leaf angle has also been thoroughly studied in model species because plant density can be significantly increased with more erect canopy. The role of GA (Shimada et al. 2006) on leaf angle determination has been confirmed via mutant studies and the manipulation of brassinosteroid (BR) gene to reduce leaf angle was sucessfully implemented to increase rice biomass yield per unit area (Morinaka et al. 2006). If markers for these and other important genes are not included in genotypic data sets, the power to completely describe the genetic architecture of a trait by GWAS could be significantly reduced, depending on the level of LD.

Therefore, combining genome-wide markers developed by high-throughput genotyping techniques with gene-specific genotypic data in GWAS will result in a more comprehensive genetic characterization of our our traits of interes.

The simultaneous analysis of phenotype-genotype associations using LD mapping and linkage mapping provides cross validation and can increase the power and resolution to dissect complex traits (Huang and Han 2014; Korte and Farlow 2013) as demonstrated in *Arabidopsis* (Brachi et al. 2010) , rice (Famoso et al. 2011) , soybean (*Glycine max* (L.) Merr.) (Zhang et al. 2014) and maize (Hung et al. 2012;Yang et al. 2013) . In sorghum, a few studies have identified QTL related to plant architecture traits, as previously described, but no GWAS has been reported using high-throughput genotyping data to investigate biomass-determinant characteristics such as leaf angle, stem circumference and others. Considering the value of combined linkage and LD mapping analysis to discover genes controlling quantitative traits, our objectives in this study were to: (i) determine the genomic regions controlling plant height, flowering time, tiller number, internode number, panicle exsertion, panicle length, seed number, stem circumference and leaf angle as biomass yield components; (ii) compare our results with previously identified QTL for those traits, if available; (iii) investigate the association between allelic variation in GA genes and our traits of interest.

Materials and Methods

Germplasm

The panel of 315 sorghum accessions used in this study has been previously described and characterized (Casa et al. 2008; Morris et al. 2013; Mantilla Perez et al. 2014). It includes 214 conversion lines and 101 historical and elite lines from public breeding programs.

Phenotypic data

Sorghum lines were planted in a randomized complete block design (RCBD) in three locations in Iowa with two replications per location, during summer 2010 and 2012. Each plot was a single 3-m-long row with 76 cm row spacing. In 2010, three representative plants per genotype per replication were evaluated in Ames, Crawfordsville and Lewis, IA for eight agronomic traits: plant height, flowering time, panicle length, panicle exsertion, stem circumference, internode number, tiller number and seed number. In 2012, two representative plants per genotype per replication were characterized for leaf angle in Ames, Crawfordsville and Greenfield, IA. Protocols implemented to measure plant height, flowering time, panicle exsertion, stem circumference and leaf angle have been previously described (Mantilla Perez et al. 2014). Internode number was determined after stripping leaves from the stem. The three panicles per genotype per replication were threshed and manually cleaned to reduce the number of small seeds that could be discarded by air blowing procedures. Counting was performed using a mechanical seed counter, and number of seeds was expressed per panicle. The number of tillers was destructively determined by manual separation from the main stalk.

Phenotypic data was analyzed by ANOVA in SAS version 9.2 (SAS Institute2008) in which location, genotype, genotype x location interaction and replication nested within location were treated as random effects. Heritability (H^2) for each trait was calculated across environments as follows:

$$H^2 = \sigma^2_G / [\sigma^2_G + (\sigma^2_{GE}/n) + (\sigma^2_e/ (nr)]$$

where σ^2_G is the genotypic variance, σ^2_{GE} is the genotype by environment interaction variance, σ^2_e is the error variance, n is the number of environments, and r is the number of

replications. Best linear unbiased prediction (BLUP) was calculated by fitting the following linear model in the R package "lme4" for the estimation of breeding values:

$$Y = (1|Genotype) + (1|Loc) + (1|RepLoc) + (1|Genotype: Loc)$$

where Y is trait data, parentheses indicate random effects, "1|" indicates random effects, and a colon (:) denotes interaction. Genotype refers to the 315 sorghum accessions, Loc refers to the three environments and Loc/Rep is replication nested within location. Correlation coefficients were calculated using BLUPs and Pearson's statistics cor procedure in R software (R core Development Team 2013).

Genotypic data

The association panel was genotyped using GBS methodology (Elshire et al. 2011). The imputed genotypic data has been previously reported (Morris et al. 2013) and is publicly available at http://www.morrislab.org/data. A total of 136,285 SNPs with minor allele frequency (MAF) > 5% and missing data < 40% were used in this study. Physical position of SNPs determined using Phytozome V1.4. A total of 263 SNPs corresponding to BR genes have been previously investigated for their potential association with plant architecture traits using the same phenotypic data set (Mantilla Perez et al. 2014) and were, thus, excluded from this study.

The Sequenom (SQNM) MassARRAY iPLEX Platform (Gabriel et al. 2009) at the Genomic Technologies Facility (Iowa State University) was used to genotype newly developed markers within GA candidate genes if no GBS data or limited number of markers were available. The identification of sorghum homologous GA genes was performed *in silico*, following a procedure similar to the one described by Mantilla Perez et al. (2014) for BR genes. In summary, previously reported GA protein sequences from model species (Chebotar and Chebotar 2010;

Davière and Achard 2013; Hedden and Thomas 2012; Yamaguchi 2008) were obtained from the National Center for Biotechnology Information databases and BLASTed against the sorghum genome sequence from phytozome V1.4 (Paterson et al. 2009) using TBLASTN. Their common domains were predicted using Pfam (Punta et al. 2012). A total of 27 GA candidate genes were identified: 19 from the biosynthesis pathway and eight from the signaling pathway (Supplemental Table S1). Twelve candidate GA genes had no markers or only one SNP from the GBS data set. Therefore, 54 new markers covering these twelve genes were developed by sequencing on ABI 3730 DNA analyzer (Applied Biosystems) and then scored using SQNM. (Supplemental Table S2). After markers were filtered for a MAF>0.05, two genes *Sb03g035000 (Gibberellin 2-oxidase)* and *Sb09g020080 (Gibberellin receptor GIDI)* did not have SNPs representing them and were thus excluded from the analysis. In summary, a total of 225 GA-related markers (from both GBS and SQNM data sets), within 25 candidate genes or 5kp upstream or downstream from them, were particularly targeted and evaluated as part of the GWAS.

To better capture the variation between 56,624,926 and 61,171,968 bp on chromosome 7, spanning the region of a previously identified QTL for leaf angle (Hart et al. 2001) were collected additional marker data. DNA from 160 accessions with imputed or missing data in the original GBS data set for the significant markers *S7_58576095* and *S7_59049004* were PCR amplified and sequenced using ABI 3730 to either confirm the imputed data or complete the GBS dataset. Missing data was reduced to < 2%. The “tandem duplication” reported as the causal polymorphism of *Dw3* (Multani et al. 2003) was genotyped in all accessions because there were no GBS SNPs available within this important gene. This was accomplished to test the hypothesis that this hormonal related gene localized within the target interval is associated with

variation inn leaf angle. In total, 683 high quality markers (MAF>0.05, and Missing data<14%) were genotyped. This data set was used to do regional single SNP association analysis in our attempt to refine the physical interval previously reported for this leaf angle QTL on chromosome 7.

Association analysis

Population structure (*Q*) for this panel has been previously estimated as five subpopulations using 702 SNPs with a minimum physical distance of 350 kb (MAF > 0.05 and missing data < 15%; Mantilla Perez et al. 2014). The same SNP data set was used to calculate the kinship matrix (*K*), an estimate of the level of relatedness among individuals, using the Loiselle algorithm (Loiselle et al. 1995) as implemented in SPAGeDi 1.4 (Hardy and Vekemans 2002).

Both general linear model (GLM including *Q*) and mixed linear model (MLM including *Q+K*) were used to test phenotype-genotype associations as implemented in TASSEL(Bradbury et al. 2007). False discovery rate (FDR), a procedure designed to control false positives as a result of multiple comparisons, was calculated using the *q* value package in R software (R core Development Team 2013).

The physical positions of previously identified QTL for our traits of interest were extracted from the following studies and the comparison with our results presented in Fig. 1: (i) plant height, flowering time, and panicle length (Hart et al. 2001; Brown et al. 2006; Srinivas et al. 2009; Zou et al. 2012; Nagaraja Reddy et al. 2013); (ii) panicle exsertion (Klein et al. 2001; Feltus et al. 2006; Brown et al. 2006); (iii) tiller number (Hart et al. 2001; Paterson et al. 1995; Feltus et al. 2006; Murray et al. 2008; Mace et al. 2009; Shiringani et al. 2010; Alam et al. 2014); (iv) internode number (Srinivas et al. 2009 and Zou et al. 2012; Nagaraja Reddy et al. 2012;

Nagaraja Reddy et al. 2013); (v) leaf angle (Hart et al. 2001); (vi) stem circumference (Zou et al. 2012); and (vii) for seed number (Nagaraja Reddy et al. 2013). Markers used in the aforementioned QTL studies were not specifically scored in this diversity panel, but their corresponding physical position was determined based on the sorghum genome sequence (Phytozome v1.4; Paterson et al. 2009) and graphically indicated in Fig. 1.

Results

Significant phenotypic variation and trait correlations

The 315 accessions used in this study exhibited a significant variation for all plant architecture traits. As previously reported (Mantilla Perez et al. 2014), genotype, location and genotype x location interaction were significant sources of variation for plant height, panicle exsertion, panicle length, stem circumference, flowering time and leaf angle. The analysis of variance also indicated that there was a significant effect of genotype, location and genotype x location interaction ($P < 0.05$) for seed number, tiller number and internode number. Detail results of the ANOVA for all traits are presented in Supplemental Table S3. The BLUPs ranged from 0.05 to 3.3 for tiller number; 6.59 to 13.88 for internode number and 387 to 3,099 for seed number. All heritability values were high (0.75-0.99) with stem circumference, tiller number and seed number being the only traits with heritability lower than 0.90 (Table 1). The mean, standard deviation and range of variation for all traits, calculated using BLUPs, are summarized in Table 1.

The correlation coefficients between all phenotypes are presented in Table 2. The strongest correlation ($r = 0.77$) was observed between flowering time and internode number. Both traits were significantly and positively correlated ($P < 0.001$) with stem circumference ($r = 0.46$ and $r = 0.57$, respectively) and seed number ($r = 0.28$ and $r = 0.41$, respectively). These

correlations suggest that as plants flowered late, they had thicker stems, more internodes and more seeds per panicle. These four traits were significantly and negatively correlated with tiller number. Plant height was positively correlated with panicle exsertion and leaf angle, while negatively correlated with stem circumference. In summary, flowering time, internode number and seed number were positively correlated with stem circumference, while stem circumference was negatively correlated with plant height, panicle exsertion and leaf angle.

Summary of genome-wide association study results

The MLM association results are presented in detail and further compared with previous knowledge of the nine traits investigated here, since this model greatly reduced the number of false positive associations when compared with GLM results, as shown in quantile-quantile plots (Supplemental Fig. S1). in the *q* value threshold was set specifically for each trait (Table 3): (i) *q* = 0.0000488 to 0.00995 (corresponding *P* value 2.6710^{-7} to 1.69×10^{-6} for leaf angle, panicle length, flowering time, tiller number, and plant height; (ii) *q* = 0.02539 to 0.06287 (corresponding P values 1.01×10^{-6} to 7.08×10^{-6}) for panicle exsertion, internode number, and seed number; and (iii) *q* = 0.1126 (corresponding P value = 4.86×10^{-5} for stem circumference. Based on these thresholds, the expected number of false positive associations was only one for leaf angle, tiller number, plant height, flowering time, panicle length, panicle exsertion, internode number and seed number and four for stem circumference. A total of 101 unique genomic regions were associated with our plant architecture traits of interest, out of 136,320 SNPs tested (Fig. 1; Table 3). Single-nucleotide polymorphisms in close physical proximity and in LD were considered part of the same significantly associated genomic region. Complete information about all significant associations identified using MLM is presented in Supplemental Table S4.

Considering that the use of a MLM could generate false negative results if causal variants are structured with kinship or between subpopulations, we identified the most significant associations obtained with GLM and compared them with MLM results. The associations uniquely identified by GLM are summarized in Supplemental Table S5.

Genome-wide association study by trait

Few markers or genomic regions with major effects were associated with phenotypic variation in plant height, leaf angle and flowering time, as shown in Table 3. For plant height, one region on chromosome 9, represented by SNP *S9_57236778*, explained 29% of the phenotype variation and another region on chromosome 6 (SNP *S6_39106643*) contributed 20% of the variation. Similar genetic architecture was observed for flowering time, since one significant SNP (*S5_51577750*) with a large effect ($R^2 = 0.157$) was identified on chromosome 5.

On chromosome 7, several SNPs spanning a 1.67 Mb region (between *S7_58178513* and *S7_59850040*), were significantly associated with variation in leaf angle, and one of them (*S7_59818811*) accounted for more than 15% of the variation (Fig. 2). The level of LD was variable within this region (Fig. 2c), but some markers were in high LD, and thus, further studies are needed to dissect this important chromosomal segment and fully understand the genes or polymorphisms controlling this phenotype.

For some traits, like stem circumference and panicle exsertion, many markers or regions with small effects were identified across several chromosomes (Table 3). Significant SNPs for variation in stem circumference were detected on all chromosomes except 6 and 9 with R^2 values that ranged from 0.055 to 0.122. Variation in panicle exsertion was explained by markers located on chromosomes 1, 2, 3, 6, 7, 9 and 10 with small effects ($0.076 < R^2 < 0.118$).

Tiller number and internode number were plant architecture traits controlled by several SNPs with relatively large effects, located on multiple chromosomes. Markers on chromosomes 1, 3, 4, 8, 9 and 10 explained 8.9% to 14.4% of the variation in tiller number. Genomic regions on chromosomes 1, 2, 5, 6, 8 and 9 controlled between 9.1% and 14.3% of the variation in internode number. A small region on chromosome 6 was the most significantly associated with variation in seed number and, five markers in that region (*S6_57048727*, *S6_57049108*, *S6_57049169*, *S6_57049184* and *S6_57049320*) correspond to polymorphisms on *KS3*, a GA biosynthetic gene similar to *Ent-kaurene synthase* (*KS*) (Supplemental Table S4).

Same single-nucleotide polymorphisms associated with different traits

In several cases, one or more SNPs were significantly associated with more than one trait, a phenomenon that could be due to pleiotropy or different causal genes in LD (Supplementary Table S4), for example: (i) SNP *S9_52325578* associated with variation in both flowering time and internode number; (ii) SNPs *S9_57836978* and *S9_58005176* explained variation in both plant height and panicle exsertion; (iii) SNPs *S6_42703814*, *S6_42726564*, *S6_42764790* were significant for both plant height and internode number; and (iv) SNP *S7_59261924* was identified associated with panicle length and stem circumference.

Association of gibberellin candidate genes with plant architecture

Two sorghum GA candidate genes, *KS3*, *Gibberellin 2- oxidase 5 (GA2ox5)*, were significantly associated with plant architecture characteristics and explained 9.1% of the variation in seed number and 14.6% of the variation in plant height, respectively. Ten markers were discovered within KS3 and genotyped by SQNM. Seven of them with high quality data (MAF>0.05, missing data< 10%) were associated with seed number variation and five of them represented the most significant genome-wide markers for the trait (Fig. 3). *GA2ox5* is a strong

candidate for further investigation because SNPs within this gene were not only significant in this study but they co-localized with previously identified QTL for plant height.

Discussion

Comparison between our genome-wide association study results and previous studies

Information for 119 previously identified QTL controlling our plant architecture traits of interest was collected and compiled into 73 narrow regions, indicated as shaded areas in Fig. 1. When compared with our GWAS results (using MLM), we observed that (i) there were 10 overlapping regions between our significant SNPs and previously identified QTL; (ii) nine significant SNPs did not fall into any previously reported QTL regions but were relatively close to them (86 kb to 2.5 Mb); (iii) three SNPs out of the nine described in (ii) were in LD with previously reported QTL and were thus not considered novel regions (Table 4); (iv) 88 significant SNPs represented newly identified regions since they neither colocalized with known QTL nor were they in LD; and (v) no significant SNPs were detected in this study within some previously identified QTL regions.

Plant height in sorghum has been extensively studied using both linkage and LD mapping. Based on this previous knowledge, *Dw1* has been mapped on chromosome 9 (Brown et al., 2008) and *Dw2* on chromosome 6 (Zou et al. 2012; Nagaraja Reddy et al. 2013). *Dw3*, localized on chromosome 7, has been cloned and the causal polymorphism has been identified as an ~800-bp tandem duplication (Multani et al. 2003). Strong association signals that correspond to *Dw1* and *Dw2* on chromosomes 9 and 6, respectively, were detected in LD mapping studies conducted on a diversity panel (Morris et al., 2013a; Zhang et al., 2015a) and on a minicore collection across multiple environments (Upadhyaya et al., 2012a). In our study, markers on chromosome 6 and 9 that correspond to *Dw1* and *Dw2* were also significantly associated with

variation in plant height, which validates our results. Additionally, Higgins et al. (2014) identified a SNP associated with variation in plant height that was very close to a *GA2ox* gene on chromosome 9 (at ~57 Mb). They, and other researchers (Brown et al. 2008; Morris et al. 2013a), suggested that this *GA2ox* gene could underlie the *Dw1* locus. Our results directly confirmed the significant association of *GA2ox5* (*Sb09g028360* at 57,265,477 bp on chromosome 9) with variation in plant height and provides additional evidence as the possible underlying gene in *Dw1* locus. However, only *GA2ox5* SNP, located on the 5'UTR (*S9_57266896*), was not the most significant marker in this chromosomal interval (Fig. 1). In spite of the quantitative genetic evidence proposing that allelic variation in *GA2ox5* is controlling plant height (Brown et al. 2008; Wang et al. 2012; Morris et al. 2013a; Thurber et al. 2013), Ordonio et al. (2014) concluded that this GA catabolic enzyme could not be *Dw1* based on two arguments. First, they indicated that if indeed *GA2ox5* was *Dw1*/*dw1*, the short phenotype should be accompanied by a bending stem, the observed response to the GA inhibitor uniconazole. However, the authors did not address the fact that *GA2ox* is encoded by a gene family in sorghum, and thus, because of the functional redundancy, an extreme bending phenotype would be unlikely. Second, expression differences for *GA2ox5* were not statistically significant between *Dw1* and *dw1* lines, but RNA sampling was only performed from elongating internodes in seedlings. It would be pertinent to test expression patterns from multiple developmental stages and tissues considering that, in rice, members of the *GA2ox* family have differential expression in various tissues (Sakamoto et al. 2004). In general, we can conclude that current knowledge of the significantly associated region on chromosome 9 suggests that *GA2ox5* is still an important candidate gene worth studying and validating.

Panicle length and flowering time have also been widely investigated by linkage mapping, and those studies, used as independent validations, provide robustness to our data. A panicle length QTL was consistently identified in the same physical interval between SNPs *S7_58285987* and *S7_6117196858* by four groups using three different biparental populations (Hart et al. 2001; Brown et al. 2006; Srinivas et al. 2009; Nagaraja Reddy et al. 2013). In our study, the same region (represented by SNP *S7_58395536*, *S7_59261924*, *S7_59503366*, and *S7_60382080*) was coincidently identified together with novel intervals on chromosome 3, 5, and 10. The region on chromosome 5 associated with variation in flowering time (*S5_51577750*) confirmed a previously reported QTL for this trait (between *S5_18472314* and *S5_55039064*; Nagaraja Reddy et al.,2013) and narrowed its confidence interval. The SNP on chromosome 10 (*S10_54425412*) significantly associated with variation in flowering time was in LD ($r_2 < 0.7$) and 82 kb away from the QTL reported by Hart et al. (2001), so we did not consider it a novel region (Table 4). New regions controlling flowering time were identified on chromosomes 8 and 9, but *Ma1*, a well-known major gene mapped to chromosome 6 (Murphy et al. 2011; Zou et al. 2012), was not associated with the trait in our study. According to Murphy et al. (2011), the *Ma1* allele delayed flowering, while the alternative *ma1* allele, present in elite lines such as BTx406, conferred earliness. The only SNP on *Ma1* tested in this study (*S6_40286721*) was not significant even though allele frequencies were intermediate. Considering the large proportion of converted and elite materials in this panel, the causal polymorphism in *Ma1* allele is likely present at low frequency or in low LD with the tested SNP, reducing the power to detect its association with variation in flowering time.

For panicle exsertion, Klein et al. (2001) detected a major QTL on chromosome 1 that explained 10.9% of the variation and was delimited by SSR markers *Xtxp37* and *Xtxp61*.

Another major QTL explaining 12.9% of the variation on chromosome 3 was identified by Feltus et al. (2006) between markers *Xtxs1175* and *Xcdo1160.* Both QTL regions, represented by SNPs *S1_64973389* and *S3_59444402*, respectively, were coincidently associated with variation in panicle exsertion in this study in addition to novel regions on chromosomes 2, 6, 7, 9, and 10. Similar results were obtained for tiller number, since one significant marker, *S1_51507363*, colocalized with a previously identified QTL between *SHO68* and *PSB062* (Paterson et al. 1995; Feltus et al. 2006; Alam et al. 2014) and another SNP (*S4_ 52740396*) corresponds to a QTL between *xtxp12* and *mscir300* (Alam et al. 2014). Newly associated regions on chromosome 3, 8, 9, and 10 were identified, as well.

No overlap between our results and known QTL regions (using biparental populations) was observed for internode number. However, a recent GWAS study investigating the genetic control of number of nodes discovered several significant SNPs in the regions 41.5 to 46.3 Mb and 42.1 to 48.7 Mb of chromosome 6. Node and internode number are, of course, two highly correlated traits, and our significant SNPs on chromosome 6 (*S6_42703814*, *S6_42726564*, *S6_42764790*, and *S6_45929612*) are localized within the same significant chromosomal interval for number of nodes reported by Zhang et al. (2015a).

In the only previous sorghum study performed to investigate the genetic control of leaf angle, a major QTL (*QLea.txs-E*) was discovered on chromosome 7 that explained 45% of the phenotypic variation in one environment and 28.4% in another environment (Hart et al. 2001). Several SNPs within this region were significant in our study with *S7_59818811* having the strongest association signal for the trait (Fig. 1, 2). In addition to this region with a major effect, other markers were significant for leaf angle on chromosomes 1, 3, 6, and 9.

Knowledge about the genetic architecture controlling stem circumference and seed number is limited in sorghum. Two QTL for stem circumference were localized on bin1535 of chromosome 4 and bin 2461 of chromosome 7 (Zou et al. 2012) under two contrasting conditions: short and long days. No markers were identified in our study within those intervals, but two significant SNPs, *S4_55448299* and *S7_59261924*, were in LD (r_2 = 0–0.59 and r_2 = 0–0.74) with the respective QTL (Table 4). Seed number, an important yield component for grain and forage sorghums, was investigated by two different groups (Brown et al. 2006; Nagaraja Reddy et al. 2013), but only one QTL was discovered on chromosome 6 between markers *gpsb069* and *Xcup12* that explained 5% of the phenotypic variance (Nagaraja Reddy et al. 2013). We have identified five significant makers (*S6_57049320*, *S6_57048727*, *S6_57049108*, *S6_57049169*, and *S6_57049184*) in the candidate GA biosynthetic gene *KS3* (*Sb06g028210*) that explained 9.1% of the phenotypic variation and that were 2.5 Mb from the previously identified QTL but not in LD (r_2 < 0.17) (Fig. 3b,c); therefore, these markers belong to a novel genomic region controlling seed number per panicle. *KS3* (*Sb06g028210*) is an interesting candidate gene for validation and further studies because it is in tandem with another KS-similar gene *KS1* (*Sb06g208220*), whose markers were not significantly associated with variation in seed number (Fig. 3b). Single-nucleotide polymorphisms representing *Sb06g028210* were not originally present in the GBS data set and in spite of the intermediate-to-high level of LD previously reported in sorghum (Hamblin et al. 2005; Morris et al. 2013a), this important genomic region on chromosome 6 would have been undetected if we had not collected additional marker data based on previous knowledge of GA genes and their effects on plant architecture

from model species. Additional novel associations were identified for both stem circumference and seed number, as indicated in Table 3 and Supplemental Table S4.

In summary, several significant markers colocalized with previously identified QTL regions for all our target traits except seed number. We recognize that some of the novel regions identified in our study could have been previously discovered, but the comparison with a few QTL studies was not possible because of limitations inherent to the marker technology used at the time, for example, amplified fragment length polymorphism and diversity array technology markers. Both the validated and novel regions reported in this study represent valuable knowledge that could be further investigated and exploited in breeding programs and significantly enrich our understanding of the genetic control of traits with limited previous information in sorghum.

Increasing resolution of a previously identified quantitative trait loci for leaf angle

Considering the importance of leaf angle for the genetic improvement of both biomass and grain yield on a per-area basis, we further investigated the significantly associated chromosomal segment that corresponds to the QTL *QLea. txs E* identified by Hart et al. (2001), and we were able to reduce the physical region to a 1.67 Mb interval. Leaf angle, defined as the inclination between leaf blade and the vertical culm (Zhao et al. 2010), is mainly determined by the joint connecting the blade with the sheath. Most mutants for leaf angle in model species had been described as having an abnormal division and expansion of adaxial cells in the collar (Nakamura et al. 2009; Zhao et al. 2010) and allelic changes in BR biosynthesis and/or signaling genes (Wada et al. 1981; Yamamuro et al. 2000; Wang et al. 2008; Tanaka et al. 2009). In sorghum, BR genes have also been associated with natural variation in leaf angle (Mantilla Perez et al. 2014), but increasing evidence suggests that other phytohormones, such as auxin, ethylene,

abscisic acid, and gibberellins, are involved in leaf angle determination as well (Cao and Chen 1995; Shimada et al. 2006; Xu et al. 2014). Detail mechanistic information for this individual group of hormones on leaf angle is not available since many of them work synergistically with BR (Cohen and Meudt 1983; Shimada et al. 2006; Hardtke et al. 2007; Song et al. 2009).

After scanning the refined genomic region on chromosome 7 for candidate genes, we detected seven that have been reported in other species as directly or indirectly affecting leaf angle (Fig. 2d; Supplemental Table S7). These seven candidate genes can be divided into two categories: hormone-related and non-hormone-related genes. In the hormone-related category, one gene, *Sb07g023360*, was associated with abscisic acid (Xu et al. 2014); two genes, *Sb07g023575* and *Sb07g023803*, were related to the ethylene pathway because of their predicted AP2 domains (Jiang et al., 2012); two genes, *Sb07g024740* and *Sb07g024750*, were involved in auxin regulation (Kant et al., 2009); and one gene, *Sb07g024110*, was related to the GA signaling pathway (Shimada et al., 2006). The only non-hormone-related gene (*Sb07g023380*) was a type of CAMK that includes calcium- and calmodulin-dependent protein kinases (Yang and Komatsu, 2000). Current knowledge of the seven novel candidates in this region indicates that *Sb07g023360* is the sorghum orthologous of *OsZHD1* gene, a zinc finger homo domain class homeobox transcription factor that plays an important role in rice morphogenesis especially in the formation and distribution of bulliform cells. Overexpression of *OsZHD1* in rice induced abaxially curled and drooping leaves (Xu et al. 2014). *Sb07g023380* is predicted to be the ortholog of a Calcium-dependent protein Kinase (CDPK) involved in the Ca^{2+}-dependent protein phosphorylation leading to brassinolide, thus affecting lamina inclination in rice (Yang and Komatsu 2000). The rice *SPINDLY* (*SPY*) gene (orthologous to *Sb07 g024110*) encodes an O-linked N-acetylglucosamine transferase considered to be a negative regulator of

GA signaling. Transgenic rice plants transformed with an *OsSPY* RNAi construct showed a larger bending angle at the lamina joint (Shimada et al. 2006). *Sb07g023575* and *Sb07g023803* were selected as candidate genes in this region for future studies because they contain an AP2 domain. An AP2 transcription factor- like gene affected internode length, leaf shape, and leaf angle in maize because of a rearrangement of leaf epidermal cells and internode parenchyma cells (Jiang et al. 2012). Both *Sb07g024740* and *Sb07g024750* belong to a SAUR family, and they were predicted to be orthologous to *SAUR36* genes involved in auxin regulation. Although there is no evidence to demonstrate that *SAUR36* functions in altering leaf angle, their family member, *SAU39*, has been verified as a key player in changing leaf angle in rice (Kant et al. 2009). Transgenic plants with single copy insertions of *SAUR39* developed more horizontal young and old leaves in 10 wk, while wild-type rice plants maintained small leaf angles.

Dw3, a well-known auxin transporter gene with a major effect on sorghum plant height, is also physically located in this important region, and it has recently been reported having pleiotropic effects on leaf angle (Truong et al. 2015). Considering that there were no GBS markers representing this gene in our genotypic data set, we specifically genotyped the association panel for the tandem repeat reported as the causal polymorphism for plant height (Multani et al. 2003). Even though we identified an interval on chromosome 7 controlling leaf angle that coincides with previously reported QTL (Supplemental Fig. S2), our results do not support the hypothesis that *Dw3* underlies variation in leaf angle (Fig. 2b). Several experimental differences between our study and Truong et al. (2015) could potentially explain these apparent contradicting results. The angle investigated in our study corresponds to the leaf immediately under the flag leaf, and it was determined at flowering time when vegetative growth ceased. Truong et al. (2015) investigated angles of the third, fourth, and fifth leaf under the leaf whorl at

several intervals before and during flowering (based on reported dates). Considering the known function of *Dw3* as an auxin transporter (Multani et al. 2003; Brown et al. 2008), the hormonal concentration would decrease from top to bottom in a plant carrying the *dw3* allele but would be homogenous throughout the stem and canopy in a *Dw3* plant. If *Dw3/dw3* is indeed controlling leaf angle, it is logical to conclude that phenotypic differences between *Dw3* and *dw3* plants would be maximized in lower leaves, in agreement with Truong et al. (2015), but not on upper leaves as suggested by our results. Therefore, we propose that additional genes in this region of chromosome 7 control leaf angle. This hypothesis is also supported by Truong et al. (2015), in which another QTL controlling leaf angle was detected close to *Dw3* in a RIL population in which both parents carried the *Dw3* allele (R07018 X R07020) (Supplemental Fig. S2). Finally, it should be acknowledged that the apparent contradicting results about the role of *Dw3* in leaf angle control could be the consequence of synthetic associations in the region, a phenomenon that has been previously described as the cause of inaccurate association signals (Dickson et al., 2010; Morris et al., 2013b; Higgins et al., 2014).

Acknowledgments

The authors appreciate Xiaochen Sun's contribution for the development of the R code for the estimation of BLUPs. This work was supported by the USDA, National Institute of Food and Agriculture [Project #IOW05298] and by the R.F. Baker Endowment, R.F. Baker Center for Plant Breeding, Iowa State University, Ames, IA 50011, USA.

References

Adeyanju, A., C. Little, J. Yu, and T. Tesso. 2015. Genome-wide association study on resistance to stalk rot disease in grain sorghum. G3: Genes, Genomes, Genet. 5:1165–1175. doi:10.1534/g3.114.016394

Alam MM, Mace ES, van Oosterom EJ, et al (2014) QTL analysis in multiple sorghum populations facilitates the dissection of the genetic physiological control of tillering. Theor Appl Genet 127:2253–2266 doi: 101007/s00122-014-2377-9

Atwell S, Huang YS, Vilhjálmsson BJ, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines Nature 465:627–631 doi: 101038/nature08800

Brachi B, Faure N, Horton M, et al (2010) Linkage association mapping of Arabidopsis thaliana flowering time in nature PLoS Genet 6:40 doi: 101371/journalpgen1000940

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: Software for association mapping of complex traits in diverse samples Bioinformatics 23:2633–2635 doi: 101093/bioinformatics/btm308

Brown PJ, Klein PE, Bortiri E, Acharya CB, Rooney WL, Kresovich S. (2006) Inheritance of inflorescence architecture in sorghum. Theor. Appl. Genet. 113:931–942. doi:10.1007/s00122-006-0352-9

Brown PJ, Rooney WL, Franks C, Kresovich S (2008) Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes Genetics 180:629–637 doi: 101534/genetics108092239

Cao H, Chen S (1995) Brassinosteroid-induced rice lamina joint inclination its relation to indole-3-acetic acid etylene Plant Growth Regul 16: 189-196

Casa AM, Pressoir G, Brown PJ, et al (2008) Community resources strategies for association mapping in Sorghum Crop Sci 48:30–40 doi: 102135/cropsci2007020080

Chandler, P.M., A. Marion-Poll, M. Ellis, and F. Gubler (2002) Mutants at the Slender1 locus of barley cv. Himalaya. Molecular and physiological characterization. Plant Physiol. 129:181–190. doi:10.1104/pp.010917

Chebotar GO, Chebotar SV (2010) Gibberellin-Signaling Pathways in Plants Cytology. Genetics 45:259-268

Cohen, J.D., and W.J. Meudt. 1983. Investigations on the mechanism of the brassinosteroid response: I. Indole-3-acetic acid metabolism and transport. Plant Physiol. 72:691–694. doi:10.1104/pp.72.3.691

Davière J-M, Achard P (2013) Gibberellin signaling in plants Development. 140:1147–51 doi: 101242/dev087650

Dickson, S.P., K. Wang, I. Krantz, H. Hakonarson, and D.B. Goldstein (2010) Rare variants create synthetic genome-wide associations. PLoS Biol. 8(1):E1000294. doi:10.1371/journal.pbio.1000294

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:1–10 doi: 101371/journalpone0019379

Famoso AN, Zhao K, Clark RT, et al (2011) Genetic Architecture of Aluminum Tolerance in Rice (Oryza sativa) Determined through Genome-Wide Association Analysis QTL Mapping. PLoS Genet 7:e1002221 doi: 101371/journalpgen1002221

Feltus FA, Hart GE, Schertz KF, et al (2006) Alignment of genetic maps QTLs between inter-intra-specific sorghum populations. Theor Appl Genet 112:1295–1305 doi: 101007/s00122-006-0232-3

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54: 357-374

Gabriel S, Ziaugra L, Tabbaa D (2009) SNP genotyping using the sequenom massARRAY iPLEX Platform. Curr Protoc Hum Genet 1–18 doi: 101002/0471142905hg0212s60

Hamblin MT., Salas Fernandez MG, Casa AM, Mitchell SE, Paterson AH, Kresovich S. 2005. Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass Sorghum bicolor. Genetics171:1247–1256.doi:10.1534/genetics.105.041566

Harberd NP, Freeling M (1989) Genetics of dominant gibberellin-insensitive dwarfism in maize. Genetics 121:827–838

Hardtke CS, Dorcey E, Osmont KS, Sibout R (2007) Phytohormone collaboration: Zooming in on auxin-brassinosteroid interactions. Trends Cell Biol 17: 485-492

Hardy, O.J., and X. Vekemans (2002) SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2:618–620. doi:10.1046/j.1471-8286.2002.00305.x

Hart GE, Schertz KF, Peng Y, Syed NH (2001) Genetic mapping of Sorghum bicolor (L) Moench QTLs that control variation in tillering other morphological characters. Theor Appl Genet 103:1232–1242 doi: 101007/s001220100582

Hedden P, Thomas SG (2012) Gibberellin biosynthesis its regulation. Biochem J 444:11–25 doi: 101042/BJ20120245

Higgins, RH, Thurber CS, Assaranurak I, and Brown PJ. (2014) Multiparental mapping of plant height and flowering time QTL in partially isogenic sorghum families. G3: Genes, Genomes, Genet. 4:1593–602

Huang X, Han B (2014) Natural variations genome-wide association studies in crop plants. Annu Rev Plant Biol 65:531–51 doi: 101146/annurev-arplant-050213-035715

Huang X, Wei X, Sang T, et al (2010) Genome-wide association studies of 14 agronomic traits in rice lraces. Nat Genet 42:961–967 doi: 101038/ng695

Hung H-Y, Shannon LM, Tian F, et al (2012) ZmCCT the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc Natl Acad Sci U S A 109:E1913–E1921 doi: 101073/pnas1203189109

Ikeda A, Ueguchi-Tanaka M, Sonoda Y, et al (2001) slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. Plant Cell 13:999–1010 doi: 101105/tpc135999

Jiang F, Guo M, Yang F, et al (2012) Mutations in an AP2 transcription factor-like gene affect internode length leaf shape in maize. PLoS One doi: 101371/journalpone0037040

Kant S, Bi YM, Zhu T, Rothstein SJ (2009) SAUR39, a small auxin-up RNA gene, acts as a negative regulator of auxin synthesis transport in rice. Plant Physiol 151:691–701 doi: 101104/pp109143875

Klein RR, Rodriguez-Herrera R, Schlueter JA, et al (2001) Identification of genomic regions that affect grain-mould incidence other traits of agronomic importance in sorghum. Theor Appl Genet 102:307–319 doi: 101007/s001220051647

Korte A, Farlow A (2013) The advantages limitations of trait analysis with GWAS: a review. Plant Methods 9:29 doi: 101186/1746-4811-9-29

Leiser, W.L., H.F. Rattunde, E. Weltzien, N. Cisse, M. Abdou, A. Diallo, A.O. Toure, J.V. Magalhaes, and B.I. Haussmann (2014) Two in one sweep: Aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West African sorghum. BMC Plant Biol. 14:206. doi:10.1186/s12870-014-0206-6

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *psychotria officinalis* (rubiaceae). Am. J. Bot 82(11):1420-1425

Lübberstedt T, Melchinger AE, Klein D, et al (1997) QTL mapping in testcrosses of European flint lines of maize: II Comparison of different testers for forage quality traits. Crop Sci 37:1913–1922 doi: 102135/cropsci19970011183X003700030037x

Mace, E.S., J.F. Rami, S. Bouchet, P.E. Klein, R.R. Klein, A. Kilian, P. Wenzl, L. Xia, K. Halloran, and D.R. Jordan (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput diversity array technology (DArT) markers. BMC Plant Biol. 9:13. doi:10.1186/1471-2229-9-13

Mantilla Perez, M.B., J. Zhao, Y. Yin, J. Hu, and M.G. Salas Fernandez (2014) Association mapping of brassinosteroid candidate genes and plant architecture in a diverse panel of Sorghum bicolor. Theor. Appl. Genet. 127:2654–2662. doi:10.1007/s00122-014-2405-9

Morinaka Y (2006) Morphological Alteration Caused by Brassinosteroid Insensitivity Increases the Biomass Grain Production of Rice. Plant Physiol 141:924–931 doi: 101104/pp106077081

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, and S. Kresovich. (2013a) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc. Natl. Acad. Sci. USA 110:453–458. doi:10.1073/pnas.1215985110

Morris, G.P., D.H. Rhodes, Z. Brenton, P. Ramu, V.M. Thayil, S. Deshpande, C.T. Hash, C. Acharya, S.E. Mitchell, E.S. Buckler, J.M. Yu, and S. Kresovich (2013b) Dissecting genome-wide association signals for loss of-function phenotypes in sorghum flavonoid

Multani DS, Briggs SP, Chamberlin MA, et al (2003) Loss of an MDR transporter in compact stalks of maize br2 sorghum dw3 mutants. Science (80-) 302:81–84 doi: 101126/science1086072

Murphy RL, Klein RR, Morishige DT, et al (2011) Coincident light clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. Proc Natl Acad Sci U S A 108:16469–16474 doi: 101073/pnas1106212108

Murray, S.C., W.L. Rooney, M.T. Hamblin, S.E. Mitchell, and S. Kresovich (2009) Sweet sorghum genetic diversity and association mapping for brix and height. Plant Genome 2:48–62. doi:10.3835/plantgenome2008.10.0011

Murray SC, Sharma A, Rooney WL, et al (2008) Genetic improvement of sorghum as a biofuel feedstock: I QTL for stem sugar grain nonstructural carbohydrates. Crop Sci 48:2165–2179 doi: 102135/cropsci2008010016

Myles S, Peiffer J, Brown PJ, et al (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202 doi: 101105/tpc109068437

Nagaraja Reddy R, Madhusudhana R, Murali Mohan S, et al (2013) Mapping QTL for grain yield other agronomic traits in post-rainy sorghum [Sorghum bicolor (L) Moench]. Theor Appl Genet 126:1921–1939 doi: 101007/s00122-013-2107-8

Nagaraja Reddy, R., R. Madhusudhana, S. Murali Mohan, D.V.N. Chakravarthi, and N. Seetharama (2012) Characterization, development and mapping of Unigene-derived microsatellite markers in sorghum [Sorghum bicolor (L.) Moench]. Mol. Breed. 29:543–564. doi:10.1007/s11032-011-9571-0

Nakamura A, Fujioka S, Takatsuto S, Tsujimoto M, et al (2009) Involvement of C-22-hydroxylated brassinosteroids in auxin-induced lamina joint bending in rice Plant. Cell Physiol 50:1627–1635

Newell MA, Asoro FG, Scott MP, et al (2012) Genome-wide association study for oat (Avena sativa L) beta-glucan concentration using germplasm of worldwide origin. Theor Appl Genet 125:1687–1696 doi: 101007/s00122-012-1945-0

Ordonio RL, Ito Y, Hatakeyama A, et al (2014) Gibberellin deficiency pleiotropically induces culm bending in sorghum: an insight into sorghum semi-dwarf breeding. Sci Rep 4:5287 doi: 101038/srep05287

Pasam RK, Sharma R, Malosetti M, et al (2012) Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol 12:16 doi: 101186/1471-2229-12-16

Paterson AH, Bowers JE, Bruggmann R, et al (2009) The Sorghum bicolor genome the diversification of grasses. Nature 457:551–556 doi: 101038/nature07723

Paterson a H, Schertz KF, Lin YR, et al (1995) The weediness of wild plants: molecular analysis of genes influencing dispersal persistence of johnsongrass, Sorghum halepense (L) Pers. Proc Natl Acad Sci U S A 92:6127–6131 doi: 101073/pnas92136127

Peng J, Carol P, Richards DE, et al (1997) The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses. Genes Dev 11:3194–3205 doi: 101101/gad11233194

Peng JR, Richards DE, Hartley NM, et al (1999) "Green revolution" genes encode mutant gibberellin response modulators. Nature 400:256–261 doi: 101038/22307

Punta M, Coggill PC, Eberhardt YR, Mistry J, et al (2012) The Pfam protein families database Nucleic Acids Res 40:D290-D301

R core Team 2013 R: A language environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0 http://wwwr-projectorg/ Accessed 05 November 2013

Rhodes, D.H., L. Jr. Hoffmann, W.L. Rooney, P. Ramu, G.P. Morris, and S. Kresovich (2014) Genome-wide association study of grain polyphenol concentrations in global sorghum [Sorghum bicolor (L.) Moench] germplasm. J. Agric. Food Chem. 62:10916–10927. doi:10.1021/jf503651t

Rooney WL, Blumenthal J, Bean B, Mullet JE (2007) Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioprod. Biorefin. 1:147–157. doi:10.1002/bbb.15

Sakamoto T, Miura K, Itoh H, Tatsumi T, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Agrawal GK, Takeda S, Abe K, Miyao A, Hirochika H, Kitano H, Ashikari M, Matsuoka M (2004) An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiol. 134:1642–1653. doi:10.1104/pp.103.033696

Salas Fernez MG, Becraft PW, Yin Y, Lübberstedt T (2009) From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci 14:454–461 doi: 101016/jtplants200906005

SAS Institute. 2008. SAS/STAT 9.2: User's guide. SAS Inst. Inc., Cary, NC.

Shehzad T, Iwata H, Okuno K (2009) Genome-wide association mapping of quantitative traits in sorghum (Sorghum bicolor (L) Moench) by using multiple models. Breed Sci 59:217–227 doi: 101270/jsbbs59217

Shimada A, Ueguchi-Tanaka M, Sakamoto T, et al (2006) The rice SPINDLY gene functions as a negative regulator of gibberellin signaling by controlling the suppressive function of the DELLA protein, SLR1, modulating brassinosteroid synthesis. Plant J 48:390–402 doi: 101111/j1365-313X200602875x

Shiringani AL, Frisch M, Friedt W (2010) Genetic mapping of QTLs for sugar-related traits in a RIL population of Sorghum bicolor L Moench. Theor Appl Genet 121:323–336 doi: 101007/s00122-010-1312-y

Song Y, You J, Xiong L (2009) Characterization of OsIAA1 gene, a member of rice Aux/IAA family involved in auxin brassinosteroid hormone responses plant morphogenesis. Plant Mol Biol 70: 297–309

Srinivas G, Satish K, Madhusudhana R, et al (2009) Identification of quantitative trait loci for agronomically important traits their association with genic-microsatellite markers in sorghum. Theor Appl Genet 118:1439–1454 doi: 101007/s00122-009-0993-6

Storey, J.D., and R. Tibshirani (2003) Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA 100:9440–9445. doi:10.1073/ pnas.1530509100

Tanaka A, Nakagawa H, Tomita C, Shimatani Z, et al (2009) BRASSINOSTEROID UPREGULATED1, encoding a helix-loop-helix protein, is a novel gene involved in brassinosteroid signaling controls bending of the lamina joint in rice Plant. Physiol 151:669–680

Thurber, CS, Ma JM, Higgins RH, Brown PJ (2013) Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol. 14: R68. doi:10.1186/gb-2013-14-6-r6

Tian F, Bradbury PJ, Brown PJ, Hung H, et al (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159-162

Truong, S.K., R.F. McCormick, W.L. Rooney, J.E. Mullet. 2015. Harnessing genetic variation in leaf angle to increase productivity of Sorghum bicolor. Genetics 201:1229–1238. doi:10.1534/genetics. 115.178608

Upadhyaya HD, Wang YH, Sharma S, et al (2012) SSR markers linked to kernel weight tiller number in sorghum identified by association mapping. Euphytica 187:401–410 doi: 101007/s10681-012-0726-9

Upadhyaya HD, Wang YH, Sharma S, Singh S, Hasenstein KH (2012b) SSR markers lined to kernel weight tiller number in sorghum identified by association mapping. Euphytica 187:401-410

Wada, K., S. Marumo, N. Ikekawa, M. Morisaki, K. Mori (1981) Brassinolide and homobrassinolide promotion of lamina inclination of rice seedlings. Plant Cell Physiol. 22:323–325

Wang, L., Y. Xu, C. Zhang, Q. Ma, S.H. Joo, S.K. Kim, Z. Xu, K. Chong (2008) OsLIC, a novel CCCH-type zinc finger protein with transcription activation, mediates rice architecture via brassinosteroids signaling. PLoS ONE 3:e3521. doi:10.1371/journal. pone.0003521

Wang, Y., P. Bible, R. Loganantharaj, H.D. Upadhyaya (2012) Identification of SSR markers associated with height using pool-based genome-wide association mapping in sorghum. Mol. Breed. 30:281–292. doi:10.1007/s11032-011-9617-3

Xu Y, Wang Y, Long Q, et al (2014) Overexpression of OsZHD1, a zinc finger homeodomain class homeobox transcription factor, induces abaxially curled drooping leaf in rice. Planta 239:803–816 doi: 101007/s00425-013-2009-7

Yamaguchi S (2008) Gibberellin Metabolism its Regulation. Annu Rev Plant Biol 59:225–251 doi: 101146/annurevarplant59032607092804

Yamamuro C, Y Ihara, X Wu, T Noguchi, S Fujioka, S Takatsuto, M Ashikari, H Kitano,M Matsuoka (2000) Loss of function of a rice brassinosteroid insensitive1 homolog prevents internode elongation and bending of the lamina joint. Plant Cell 12:1591–1606. doi:10.1105/ tpc.12.9.1591

Yang G, Komatsu S (2000) Involvement of calcium-dependent protein kinase in rice (Oryza sativa L) lamina inclination caused by brassinolide. Plant Cell Physiol 41:1243–50 doi: 101093/pcp/pcd050

Yang Q, Li Z, Li W, et al (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity accelerated the postdomestication spread of maize. Proc Natl Acad Sci 110:16969–16974 doi: 101073/pnas1310949110

Yuan JS, Tiller KH, Al-Ahmad H, et al (2008) Plants to power: bioenergy to fuel the future. Trends Plant Sci 13:421–429 doi: 101016/jtplants200806001

Zhang D, Kong WQ, Robertson J, Goff VH, Epps E, Kerr A, Mills G, Cromwell J, Lugin Y, Phillips C, Paterson A (2015a) Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). BMC Plant Biol. 15:107. doi:10.1186/s12870-015-0477-6

Zhang D, Li J, Compton RO, Robertson J, Goff VH, Epps E, Kong W, Kim C, Paterson AH (2015b) Comparative genetics of seed size traits in divergent cereal lineages represented by sorghum (Panicoidae) and rice (Oryzoidae). G3: Genes, Genomes, Genet. 5:1117–1128.doi:10.1534/g3.115.017590

Zhang D, Song H, Cheng H, et al (2014) The Acid Phosphatase-Encoding Gene GmACP1 Contributes to Soybean Tolerance to Low-Phosphorus Stress. PLoS Genet 10:8–10 doi: 101371/journalpgen1004061

Zhang, Z., Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, BucklerES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360. doi:10.1038/ng.546

Zhao J, Mantilla Perez MB, Hu J, Salas Fernez MG (2016) Genome-Wide Association Study for Nine Plant Architecture Traits in Sorghum. Plant Genome 9:0 doi:103835/plantgenome2015060044

Zhao SQ, Hu J, Guo LB, Qian Q, Xue HW (2010) Rice leaf inclination2, a VIN3-like protein, regulates leaf angle through modu-lating cell division of the collar. Cell Res 20: 935–947

Zou G, Zhai G, Feng Q, Yan S, et al (2012) Identification of QTL for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods. J Exp Bot 63:5454-5462

Figures

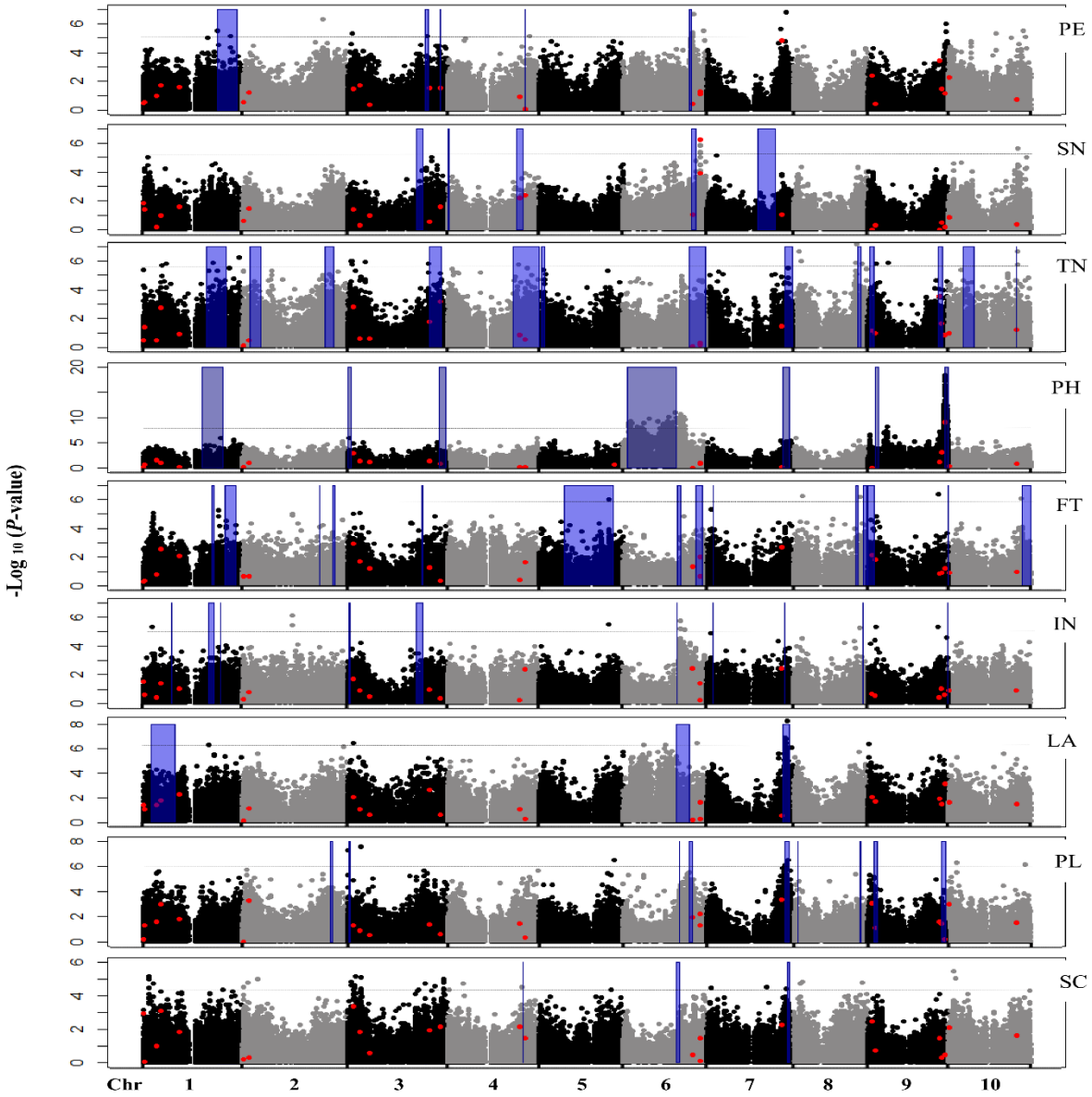


Figure 1. Genome wide association study (GWAS) results for nine plant architecture traits using 315 diverse sorghum accessions. PE, panicle exsertion; FT, flowering time; LA, leaf angle; SC, stem circumference; PH, plant height; SN, seed number; IN, internode number; TN, tiller number; PL, panicle length; shaded blue regions represent previously identified quantitative trait loci. Red dots indicate the physical position of gibberellin candidate genes. Horizontal black dotted line indicates significance threshold. Each single-nucleotide polymorphism is represented by a dot, whose center indicates the exact physical position of the marker.

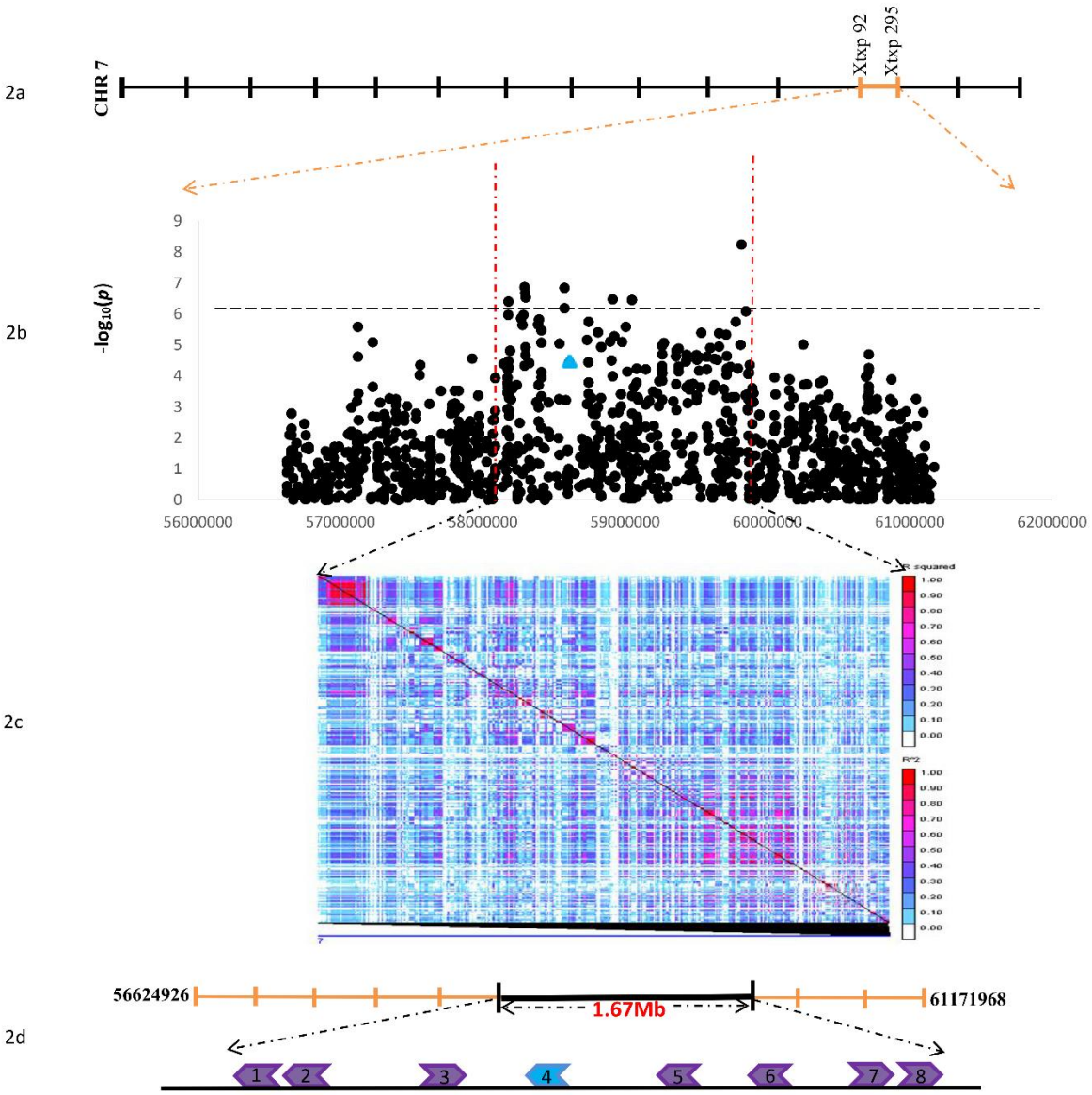


Figure 2. Increased resolution of a previously identified quantitative trai loci (QTL) for leaf angle. (a) Leaf angle locus (QTL) QLea.txs-E was previously mapped between simple-sequence repeat markers *Xtxp92* and *Xtxp295* on chromosome 7 (Hart et al. 2001). (b) narrower region of 1.67 Mb on chromosome 7 significantly associated with leaf angle; blue triangle represents the position and association significance *Dw3* (tandem duplication as function polymorphism for plant height was scored as marker). (c) Linkage Disequilibrium plot of markers within 1.67 Mb region. (d) Candidate genes within the 1.67 Mb region are indicated with colored arrows and ordered based on physical map.1: *Sb07g023360* (ZF-HD homeobox); 2: *Sb07g023380* (Kinase); 3: *Sb07g023575* (AP2 domain); 4: *Sb07g023730* (*Dw3*)(blue); 5: *Sb07g023803* (AP2 domain); 6: *Sb07g024110* (similar to *SPINDLY*); 7: *Sb07g024740* (similar to *SAUR36*); 8: *Sb07g024750* (similar to *SAUR36*).

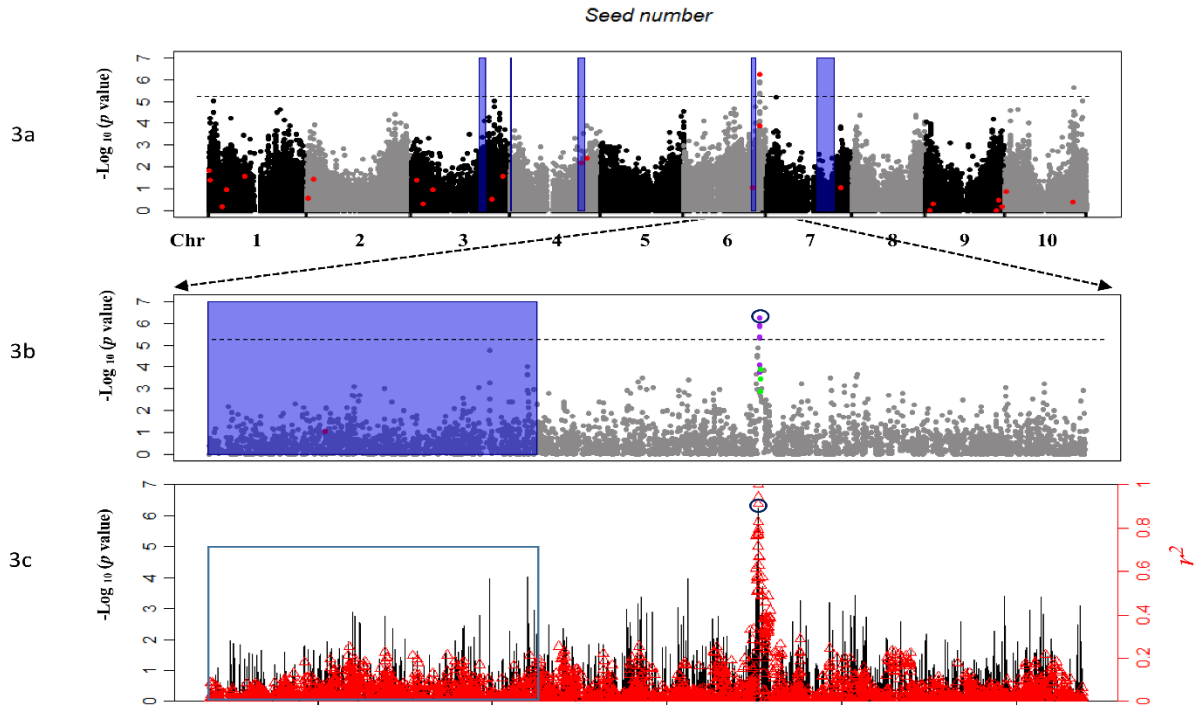


Figure 3. Close-up view of the significantly associated region on chromosome 6 affecting seed number. (a) Genome-wide association study for seed number; only one representative single-nucleotide polymorphism (SNP) for each gibberellin gene was highlighted in red to simplify the graphic representation. Shaded blue regions represent previously identified quantitative trait loci. (b) Close-up view of genomic region on chromosome 6 significantly associated with seed number in this study including the QTL (shaded blue region) reported by Nagaraja Reddy et al. (2013). Markers within two gibberellin candidate genes similar to *Ent-kaurene synthase (KS)* are colored differently than in (a): *KS3 (Sb06g028210)* with purple dots and *KS1 (Sb06g028220)* with green dots The -$\log_{10}$ *P* value for the most significant SNP (*S6_57049320)* is indicated with a purple circle. Red dot within already-know QTL region corresponds to the same nonsignificant gibberellin marker indicated in (a). (c) Scatter plot of association results and linkage disequilibrium estimates (r^2) between every SNP and the marker most significantly associated with seed number *(Sb06_57049320*) with *KS3* (*Sb06g028210*). The black vertical lines are -$\log_{10}$ *P* value for SNPs within this region; the -$\log_{10}$ *P* values for the most significant SNP (*S6_57049320)* is indicated with a blue circle. Red triangles are the r2 values calculated between each SNP and *S6_57049320.*

Tables

Table 1. Phenotypic variation of all traits based on best linear unbiased predictions (calculated as genotype performance across environments).

Traits	Units	Mean ± SD	Range	Heritability‡
PH†	cm	153.55 ± 58.36	68.56-365.91	0.99
PL†	cm	25.70 ± 6.15	9.95-55.84	0.98
PE†	cm	10.35 ± 7.63	0.02-39.03	0.95
SC†	cm	5.78 ± 0.84	3.42-8.26	0.88
TN	number	0.65 ± 0.56	0.05-3.3	0.75
IN	number	10.65 ± 1.34	6.59-13.88	0.92
FT†	day	67.46 ± 3.94	54.45-77.16	0.94
LA†	degree	50.52 ± 13.31	12.92-88.64	0.95
SN	number	1567.63 ± 502.9	387-3099	0.88

PH, plant height; PL, panicle length; PE, panicle exsertion; SC, stem circumference; TN, tiller number; IN, internode number; FT, flowering time; LA, leaf angle; SN, seed number; †Corresponding traits were also previously reported (Mantilla Perez et al. 2014);‡$H^2 = \sigma^2_G / [\sigma^2_G + (\sigma^2_{GE}/n) + (\sigma^2_e/(nr)]$, where σ^2_G is the genotypic variance, σ^2_{GE} is the genotype x environment interaction variance, σ^2_e is the error variance, n is the number of environments, and r is the number of replications.

Table 2. Phenotypic correlations between traits based on best linear unbiased predictions.

	Correlation (*r*)								
Traits	PH	PL	PE	SC	TN	IN	FT	LA	SN
PH†	–								
PL†	0.15	–							
PE†	0.47**	0.11	–						
SC†	-0.30**	0.10	-0.22**	–					
TN	-0.01	0.02	0.07	-0.46**	–				
IN	0.19*	0.01	-0.09	0.57**	-0.47**	–			
FT†	0.15	0.16*	-0.10	0.46**	-0.31**	0.77**	–		
LA†	0.30**	-0.08	0.03	-0.20*	0.06	-0.13	-0.22**	–	
SN	-0.14	-0.14	-0.28**	0.49**	-0.37**	0.41**	0.28**	-0.16*	–

*Significant at $P < 0.01$ probability level. ** Significant at $P < 0.001$ probability level. PH, plant height; PL, panicle length; PE, panicle exsertion; SC, stem circumference; TN, tiller number; IN, internode number; FT, flowering time; LA, leaf angle; SN, seed number; †Corresponding traits previously reported (Mantilla Perez et al. 2014).

Table 3. Summary of significant single-nucleotide polymorphisms for the nine plant architecture traits

Traits	FDR Threshold	Corresponding *P*-value	Chromosome	R^2 range	No. of significant SNPs representative regions[†]
PH	$q \leq 0.0000488$	$P \leq 2.67\times10^{-7}$	6,9	0.1-0.290	6
LA	$q \leq 0.003982$	$P \leq 6.42\times10^{-7}$	1,3,6,7,9	0.092-0.159	7
PL	$q \leq 0.00995$	$P \leq 1.69\times10^{-6}$	3,5,7,10	0.091-0.139	6
TN	$q \leq 0.00773$	$P \leq 1.77\times10^{-6}$	1,3,4,8,9,10	0.089-0.144	17
FT	$q \leq 0.02539$	$P \leq 1.01\times10^{-6}$	5,8,9,10	0.123-0.157	5
IN	$q \leq 0.06287$	$P \leq 7.08\times10^{-6}$	1,2,5,6,8,9	0.091-0.134	8
SN	$q \leq 0.04597$	$P \leq 8.82\times10^{-6}$	6,10	0.08-0.091	2
PE	$q \leq 0.03566$	$P \leq 7.27\times10^{-6}$	1,2,3,6,7,9,10	0.076-0.118	14
SC	$q \leq 0.1126$	$P \leq 4.86\times10^{-5}$	1,2,3,4,5,7,8,10	0.055-0.122	36

PH, plant height; LA, leaf angle; PL, panicle length; TN, tiller number; FT, flowering time; IN, internode number; SN, seed number; PE, panicle exsertion; SC, stem circumference. [†]A single region is defined by SNPs in physical proximity and in LD.

Table 4. Linkage disequilibrium (LD) analysis between significant single-nucleotide polymorphisms (SNPs) and previously identified quantitative trait loci (QTL). This analysis was performed to determine the novelty of regions identified in this study.

Trait	Chr	Previously identified QTL		SNP	Distance (bp)[‡]	LD (r^2)[‡]	Comments
		Begin	End				
FT§	8	46135880	48203322	*S8_49721204*	1,517,882	0-0.2	Novel region
FT¶	10	54507175	60949262	*S10_54425412*	81,763	0-0.7	Overlap
IN#	6	40250000	40650000	*S6_42703814*	2,053,814	0-0.1	Novel region
IN#	8	52050000	52150000	*S8_49721204*	2,328,796	0-0.1	Novel region
PE#	6	51050130	51148391	*S6_52243597*	1,095,206	0-0.1	Novel region
PL††	3	1479755	1992880	*S3_788281*	691,474	0-0.15	Novel region
SC#	4	55850000	55950000	*S4_55448299*	401,701	0-0.59	Overlap
SC#	7	59450000	59550000	*S7_59261924*	188,076	0-0.74	Overlap
SN††	6	50761007	54510676	*S6_57049320*	2,538,644	0-0.17	Novel region

FT, flowering time; IN, internode number; PE, panicle exsertion; PL, panicle length; SC, stem circumference; SN, seed number.

‡ Between newly associated SNPs and previously identified QTL. The superscript following each trait indicates the QTL study.

§ Srinivas et al. (2009).

¶ Hart et al. (2001).

Zou et al. (2012).

†† Nagaraja Reddy et al. (2013). Physical positions as indicated in Phytozome v1.4.

Supplemental Data for Chapter 2

Supplemental Table S1. Identification of Gibberellin (GA) candidate genes in sorghum

Gene	Name Function	ID (Model species)	ID Sorghum	Coordinate	BLAST Identity	BLAST E-Value	Pfam Domains
CPS1‡¶	*Ent-copalyl diphosphate synthase*	*At4g02780* *Os02g0278700*	*Sb01g021990*†	26915971 - 26925014	46% 65%	0 0	Terpene synth Terpene synth C
GAMT1‡	*Gibberellic acid methyltransferase 1*	*At4g26420*	*Sb03g032230*	60648950 - 60652225	30%	6.00E-50	Methyltransf 7
KS1‡¶	*Ent-kaurene synthase*	*At1g79460* *Os04g0611800*	*Sb06g028220*†	57054005-57058330	46% 70%	0 0	Terpene synth Terpene synth C
KS3‡¶	*Ent-kaurene synthase*	*At1g79460* *Os04g0611800*	*Sb06g028210*†	57045616 - 57050067	48% 68%	3E-138 0	
CYP714D1‡	*Cytochrome P450714D1*	*Os05g0482400*	*Sb09g023680*	53324808 - 53333470	78%	0	P450
KAO1‡¶	*ent-kaurenoic acid oxidase*	*At2g32440* *Os06g0110000*	*Sb10g000920*	612730 - 616375	58% 82%	0 0	P450
KO1‡¶	*Ent-kaurene oxidase*	*At5g25900* *Os06g0570100*	*Sb10g022520*†	50447232 - 50452777	56% 80%	0 0	P450
GA20ox1‡¶	*Gibberellin 20-oxidase*	*At4g25420* *Os03g0856700*	*Sb01g000580*	462325 - 464149	59% 74%	2.00E-156 0	DIOX_N 2OG-FeII_Oxy

Supplemental Table S1. continued

GA20ox2‡¶	*Gibberellin 20-oxidase*	*At1g44090* *Os01g0883800*	*Sb03g041900*†	69376586 - 69378900	51% 83%	3.00E-121 0	DIOX_N 2OG-FeII_Oxy
GA20ox3‡¶	*Gibberellin 20-oxidase*	*Os07g0169700*	*Sb02g003940*	4376629-4377866	61%	5.00E-148	DIOX_N 2OG-FeII_Oxy
GA3ox‡	*Gibberellin 3-oxidase*	*At4g21690*	*Sb01g000650*	580090 - 581169	29%	3.00E-45	DIOX_N 2OG-FeII_Oxy
GA3ox2‡¶	*Gibberellin 3-oxidase*	*At1g15550* *Os01g0177400*	*Sb03g004020*	4284218 - 4286411	40% 81%	8.00E-88 0	DIOX_N 2OG-FeII_Oxy
GA2ox1‡¶	*Gibberellin 2-oxidase*	*At1g02400* *Os05g0158600*	*Sb09g004520*†	5485906 - 5490693	47% 73%	6.00E-95 2.00E-152	DIOX_N 2OG-FeII_Oxy
GA2ox2‡¶	*Gibberellin 2-oxidase*	*At1g47990*	*Sb03g013450*	16764890-16770701	47%	6.00E-89	DIOX_N 2OG-FeII_Oxy
GA2ox5‡¶	*Gibberellin 2-oxidase*	*At1g78440*	*Sb09g028360*	57265477 - 57266870	49%	6.00E-103	DIOX_N 2OG-FeII_Oxy
GA2ox6‡¶	*Gibberellin 2-oxidase*	*At2g34555* *Os01g0757200*	*Sb03g035000*	63174161 - 63175793	49% 83%	8.00E-112 0	DIOX_N 2OG-FeII_Oxy
GA2ox7‡¶	*Gibberellin 2-oxidase*	*Os05g0514600*	*Sb09g025470*	54884503-54885823	74%	1.00E-49	DIOX_N 2OG-FeII_Oxy
GA2ox8‡¶	*Gibberellin 2-oxidase*	*At1g50960* *Os07g0103500*	*Sb02g000460*	417567 - 418643	41% 62%	3.00E-73 3.00E-146	DIOX_N 2OG-FeII_Oxy

Supplemental Table S1. continued

GA2ox10‡ ¶	*Gibberellin 2-oxidase*	*At4g21200* *Os04g0522500*	*Sb06g022880*	52091261-52094279	49% 74%	2.00E-108 9.00E-171	DIOX_N 2OG-FeII_Oxy
SLN1§	*DELLA protein*	*18254373*#	*Sb01g010660*	9419975 - 9422628	83%	0	DELLA GRAS
RHT1§		*75207630*††			83%	0	
GAI §		*At1g14920*			55%	0	
SLR1§ ¶		*Os03g0707600*			84%	0	
GID1§ ¶	*Gibberellin receptor GID1*	*At3g05120* *Os05g0407500*	*Sb09g020080*	49162045 - 49164861	61% 83%	2.00E-155 0	Abhydrolase 3
GID2a§ ¶	*F-box protein GID2*	*At4g24210* *Os02g0580300*	*Sb04g024040*†	53756926 - 53757936	42% 61%	3.00E-30 3.00E-64	F-box
RBX1§	*RING-box protein 1*	*At5G20570*	*Sb03g008660*†	9288095 - 9290599	97%	8.00E-68	Zf-rbx1
CUL1§	*Cullin-1*	*At4g02570*	*Sb04g027970*†	57920345 - 57923199	34%	9.00E-139	Cullin Cullin _Nedd8
SKP1§	*UFO-binding protein 1*	*At1g75950*	*Sb07g021450*†	55486464 - 55486970	72%	2.00E-73	Skp poz Skp1
pIF3§	*Phytochrome-associated protein 3*	*At1g09530*	*Sb09g003090*	3387511 - 3389459	59%	8.00E-43	HLH
PIF4§	*Phytochrome-associated protein 4*	*At2g43010*	*Sb01g013843*	12946309 - 12953676	34%	8.00E-49	HLH

† GA candidate Genes in Sorghum bicolor for which limited GBS data was available and that were successfully scored using Sequenome technology.

‡ GA genes in biosynthesis pathway; § GA genes in signaling pathway; ¶ GA genes also identified by Ordonio et al., 2014; # GA protein ID in *Triticum aestivumg*

††GA protein ID in *Hordeum vulgare*; Os, *Oryza sativa*; At, *Arabidopsis thaliana*

Supplemental Table S2. Allelic information for newly developed markers on Gibberellin (GA) genes using Sequenom technology

This information can be found in supplementary information in paper

Zhao J, Mantilla Perez MB, Hu J, Salas Fernandez MG (2016) Genome-Wide Association Study for Nine Plant Architecture Traits in Sorghum. Plant Genome 9:0. doi: 10.3835/plantgenome2015.06.004

Supplemental Table S3. Analysis of variance for all traits

Trait	Source	DF	*F* Value	*P* > F
Internode Number	Loc	2	365.69	<.0001
	rep(Loc)	3	5.75	0.0007
	Genotype	314	47.3	<.0001
	Loc*Genotype	598	3.85	<.0001
Seed Number	Loc	2	13.53	<.0001
	rep(Loc)	3	3.12	0.0255
	Genotype	312	19.57	<.0001
	Loc*Genotype	599	2.45	<.0001
Tiller Number	Loc	2	61.88	<.0001
	rep(Loc)	3	2.26	0.0797
	Genotype	314	7.79	<.0001
	Loc*Genotype	592	2	<.0001
Plant Height	Loc	2	368.11	<.0001
	rep(Loc)	3	1.56	0.1969
	Genotype	314	307.33	<.0001
	Loc*Genotype	599	3.38	<.0001
Panicle Exsertion	Loc	2	33.16	<.0001
	rep(Loc)	3	0.79	0.4974
	Genotype	314	41.79	<.0001
	Loc*Genotype	593	2.4	<.0001
Panicle Length	Loc	2	54.1	<.0001
	rep(Loc)	3	1.89	0.1299
	Genotype	314	99.21	<.0001
	Loc*Genotype	599	2.16	<.0001
Stem Circumference	Loc	2	48.27	<.0001
	rep(Loc)	3	4.19	0.0059
	Genotype	314	15.77	<.0001
	Loc*Genotype	598	2.06	<.0001
Flowering Time	Loc	2	458.57	<.0001
	rep(Loc)	3	0.89	0.4481
	Genotype	271	25.37	<.0001
	Loc*Genotype	528	1.81	<.0001
Leaf Angle	Loc	2	5.96	0.0027
	rep(Loc)	3	1.64	0.1786
	Genotype	307	22.47	<.0001
	Loc*Genotype	601	1.14	0.041

Supplemental Table S4. Summary of GWAS results for all traits from MLM

Trait	Chr	Marker	R^2	*q* value	*p* value	PIQTL Region		Reference
						Begin	End	
Plant Height	6	*S6_39106643*	0.203	1.26×10^{-8}	7.71×10^{-12}	3173639	44975172	Brown et al., 2006 Nagaraja Reddy et al., 2013
		S6_31120662	0.117	4.88×10^{-5}	2.67×10^{-7}			
	9	*S9_14291470*	0.139	2.57×10^{-6}	6.51×10^{-9}	NR		
		S9_57175732	0.225	8.25×10^{-10}	2.09×10^{-13}	57140000	57210000	Brown et al., 2008
		S9_57236778	0.290	4.64×10^{-14}	3.79×10^{-19}	57219361	57656063	Nagaraja Reddy et al., 2013
		S9_57836978	0.161	4.94×10^{-8}	4.2×10^{-11}	NR		
		S9_58005176	0.192	2.84×10^{-9}	9.98×10^{-13}	NR		
Leaf Angle	1	*S1_48277144*	0.111	3.09×10^{-3}	4.73×10^{-7}	NR		
	3	*S3_4342278*	0.108	3.09×10^{-3}	3.23×10^{-7}	NR		
	6	*S6_15123197*	0.092	3.09×10^{-3}	4.76×10^{-7}	NR		
		S6_31100085	0.106	3.09×10^{-3}	4.48×10^{-7}	NR		
		S6_54776546	0.100	3.09×10^{-3}	3.29×10^{-7}	NR		
	7	*S7_58181673*	0.099	3.09×10^{-3}	3.93×10^{-7}	56624926	61171968	Hart et al., 2001
		S7_58576095	0.081	3.98×10^{-3}	6.4×10^{-7}			
		S7_59818811	0.152	2.98×10^{-4}	5.72×10^{-9}			
	9	*S9_864973*	0.110	3.09×10^{-3}	3.92×10^{-7}	NR		
Panicle Length	3	*S3_788281*	0.125	2.43×10^{-3}	4.85×10^{-8}	NR		
		S3_9911168	0.109	2.43×10^{-3}	2.95×10^{-8}	NR		
	5	*S5_55647865*	0.120	6.62×10^{-3}	3.12×10^{-7}	NR		
	7	*S7_58395536*	0.117	8.73×10^{-3}	8.72×10^{-7}	58285987	61171968	Borwn et al., 2006 Hart et al., 2001 Srinivas et al., 2009 Nagaraja Reddy et al., 2013
		S7_59261924	0.091	9.95×10^{-3}	1.69×10^{-6}			
		S7_59503360	0.139	6.99×10^{-3}	4.19×10^{-7}			
		S7_60382080	0.110	6.61×10^{-3}	3.30×10^{-7}			
	10	*S10_6013765*	0.094	7.04×10^{-3}	4.92×10^{-7}	NR		
		S10_57621898	0.093	8.17×10^{-3}	6.53×10^{-7}	NR		

Supplemental Table S4. Continued

Tiller Number	1	*S1_17311466*	0.090	7.77×10^{-3}	1.55×10^{-6}	NR		
		S1_51507363	0.090	7.77×10^{-3}	1.30×10^{-6}	46755967	53605463	Adam et al., 2014 Feltus et al., 2006
		S1_71203776	0.107	7.77×10^{-3}	5.87×10^{-7}	NR		
	3	*S3_3639209*	0.098	7.77×10^{-3}	9.61×10^{-7}	NR		
		S3_3096177	0.091	7.77×10^{-3}	9.63×10^{-7}	NR		
		S3_13487377	0.095	7.77×10^{-3}	1.12×10^{-6}	NR		
		S3_60076113	0.139	7.77×10^{-3}	1.42×10^{-6}	NR		
	4	*S4_52740396*	0.112	7.77×10^{-3}	1.77×10^{-6}	48576873	58285987	Adam et al., 2014
	8	*S8_4234625*	0.116	7.62×10^{-3}	2.12×10^{-7}	NR		
		S8_5681713	0.091	7.77×10^{-3}	1.02×10^{-6}	NR		
		S8_46335653	0.089	7.77×10^{-3}	1.28×10^{-6}	NR		
		S8_47476046	0.118	7.04×10^{-3}	6.68×10^{-8}	NR		
		S8_51386455	0.130	7.77×10^{-3}	1.61×10^{-6}	NR		
	9	*S9_6448226*	0.092	7.77×10^{-3}	1.60×10^{-6}	NR		
		S9_15312879	0.139	7.77×10^{-3}	1.34×10^{-6}	NR		
	10	*S10_51271757*	0.142	7.77×10^{-3}	1.67×10^{-6}	NR		
		S10_51495808	0.144	7.62×10^{-3}	2.17×10^{-7}	NR		
Flowering Time	5	*S5_51577750*	0.157	2.54×10^{-2}	1.01×10^{-6}	18472314	55039064	Nagaraja Reddy et al., 2013
	8	*S8_6907738*	0.123	2.54×10^{-2}	5.92×10^{-7}	NR		
		S8_49721204	0.123	2.54×10^{-2}	6.76×10^{-7}	NR		
	9	*S9_52325578*	0.126	2.54×10^{-2}	4.20×10^{-7}	NR		
	10	*S10_54425412*	0.130	2.54×10^{-2}	8.94×10^{-7}	54507175	60949262	LD with [Hart et al., 2001]
Seed Number	6	*S6_57048727*	0.091	3.80×10^{-2}	1.32×10^{-6}			
		S6_57049108	0.081	3.80×10^{-2}	1.45×10^{-6}			
		S6_57049169	0.078	4.60×10^{-2}	6.82×10^{-6}	NR		
		S6_57049184	0.080	4.60×10^{-2}	6.17×10^{-6}			
		S6_57049320	0.080	2.70×10^{-2}	5.95×10^{-7}			

Supplemental Table S4. Continued

Seedmber Number	10	*S10_51660432*	0.096	3.80×10^{-2}	1.67×10^{-6}	NR		
Internode Number	1	*S1_6300266*	0.099	6.29×10^{-2}	4.46×10^{-6}	NR		
	2	*S2_36682369*	0.121	6.29×10^{-2}	7.76×10^{-7}	NR		
	5	*S5_51577750*	0.134	6.29×10^{-2}	3.10×10^{-6}	NR		
	6	*S6_42703814*	0.091	6.29×10^{-2}	5.98×10^{-6}	NR		
		S6_42726564	0.096	6.29×10^{-2}	6.47×10^{-6}			
		S6_42764790	0.102	6.29×10^{-2}	1.69×10^{-6}			
		S6_45929612	0.143	6.29×10^{-2}	7.08×10^{-6}	NR		
	8	*S8_49721204*	0.102	6.29×10^{-2}	5.27×10^{-6}	NR		
	9	*S9_5964485*	0.104	6.29×10^{-2}	4.69×10^{-6}	NR		
		S9_52325578	0.100	6.29×10^{-2}	4.94×10^{-6}	NR		
Panicle Exsertion	1	*S1_54743734*	0.100	3.11×10^{-2}	3.26×10^{-6}	NR		
		S1_64973389	0.089	3.50×10^{-2}	6.62×10^{-6}	55123963	66750018	Klein et al., 2001
	2	*S2_59616119*	0.106	1.59×10^{-2}	5.05×10^{-7}	NR		
	3	*S3_3578141*	0.076	3.22×10^{-2}	4.57×10^{-6}	NR		
		S3_59444402	0.103	3.57×10^{-2}	7.27×10^{-6}	57927181	60207298	Feltus et al., 2006
	6	*S6_51565039*	0.086	3.13×10^{-2}	4.15×10^{-6}	NR		
		S6_52243597	0.116	1.13×10^{-2}	2.13×10^{-7}	NR		
		S6_60117699	0.090	3.11×10^{-2}	3.21×10^{-6}	NR		
	7	*S7_55229509*	0.094	2.93×10^{-2}	2.22×10^{-6}	NR		
		S7_58733063	0.118	1.13×10^{-2}	1.49×10^{-7}	NR		
	9	*S9_57836978*	0.084	3.11×10^{-2}	3.53×10^{-6}	NR		
		S9_58005176	0.097	1.59×10^{-2}	9.81×10^{-7}	NR		
	10	*S10_35739185*	0.099	3.29×10^{-2}	4.98×10^{-6}	NR		
		S10_55333956	0.089	3.11×10^{-2}	3.17×10^{-6}	NR		

Supplemental Table S4. Continued

Stem Circumference	1	*S1_4086776*	0.080	1.13×10^{-1}	1.06×10^{-5}	NR		
		S1_4119134	0.065	1.13×10^{-1}	3.23×10^{-5}	NR		
		S1_7130136	0.082	1.13×10^{-1}	1.91×10^{-5}	NR		
	1	*S1_58977109*	0.065	1.13×10^{-1}	4.19×10^{-5}	NR		
	2	*S2_739664*	0.082	1.13×10^{-1}	3.00×10^{-5}	NR		
		S2_3757857	0.073	1.13×10^{-1}	1.58×10^{-5}	NR		
	3	*S3_2451843*	0.070	1.13×10^{-1}	1.45×10^{-5}	NR		
		S3_2472883	0.067	1.13×10^{-1}	2.27×10^{-5}	NR		
		S3_4294991	0.058	1.13×10^{-1}	4.17×10^{-5}	NR		
		S3_4390904	0.055	1.13×10^{-1}	4.59×10^{-5}	NR		
		S3_4969227	0.072	1.13×10^{-1}	2.46×10^{-5}	NR		
		S3_10032457	0.082	1.13×10^{-1}	4.70×10^{-5}	NR		
		S3_10079727	0.062	1.13×10^{-1}	4.16×10^{-5}	NR		
		S3_10210662	0.073	1.13×10^{-1}	4.86×10^{-5}	NR		
		S3_10238627	0.081	1.13×10^{-1}	8.11×10^{-6}	NR		
		S3_10263132	0.071	1.13×10^{-1}	3.87×10^{-5}	NR		
		S3_53254054	0.069	1.13×10^{-1}	4.22×10^{-5}	NR		
		S3_57496346	0.091	1.13×10^{-1}	1.85×10^{-5}	NR		
		S3_62746406	0.064	1.13×10^{-1}	4.84×10^{-5}	NR		
		S3_64700033	0.079	1.13×10^{-1}	2.59×10^{-5}	NR		
	3	*S3_71474582*	0.076	1.13×10^{-1}	1.02×10^{-5}	NR		
		S3_71533774	0.092	1.13×10^{-1}	1.76×10^{-5}	NR		
		S3_71864054	0.071	1.13×10^{-1}	1.51×10^{-5}	NR		
	4	*S4_1497814*	0.102	1.13×10^{-1}	4.27×10^{-5}	NR		
		S4_11552920	0.122	1.13×10^{-1}	1.94×10^{-5}	NR		
		S4_55448299	0.065	1.13×10^{-1}	2.96×10^{-5}	55850000	55950000	LD with [Zou et al., 2012]
	5	*S5_53680290*	0.100	1.13×10^{-1}	4.49×10^{-5}	NR		

Supplemental Table S4. Continued

Stem Circumference	7	*S7_2948882*	0.087	1.13×10^{-1}	3.28×10^{-5}		NR	
		S7_44723826	0.079	1.13×10^{-1}	3.01×10^{-5}		NR	
		S7_59261924	0.067	1.13×10^{-1}	4.06×10^{-5}	59450000	59550000	LD with [Zou et al., 2012]
	8	*S8_5330018*	0.080	1.13×10^{-1}	1.84×10^{-5}		NR	
	8	*S8_6637105*	0.079	1.13×10^{-1}	1.50×10-5		NR	
	8	*S8_51210480*	0.098	1.13×10^{-1}	1.68×10^{-5}		NR	
	10	*S10_3692625*	0.082	1.13×10^{-1}	3.29×10^{-6}		NR	
		S10_4557334	0.079	1.13×10^{-1}	1.61×10^{-5}		NR	
		S10_60905954	0.064	1.13×10^{-1}	4.71×10^{-5}		NR	

Markers/regions overlapping with previously identified QTLs are highlighted in yellow.
Markers/regions outside of PIQTL regions but in LD with them are indicated in purple.
NR indicates newly identified SNP/region.

Supplementary Table S5. Complementary to MLM GWAS result from general linear model (GLM).

Trait	Chr	*Marker*	R^2	*q* value	*p* value	Previously identified QTL Region		Reference
						Begin	End	
Panicle Excertion	6	*S6_50076783*	0.096	1.13×10^{-3}	1.96×10^{-07}	NR		
	7	*S7_55192577*	0.113	5.07×10^{-4}	1.75×10^{-08}	NR		
		S7_55484074	0.100	1.1×10^{-3}	1.79×10^{-07}	NR		
	8	*S8_44384377*	0.094	1.1×10^{-3}	1.33×10^{-07}	NR		
	9	*S9_57957443*	0.105	1.1×10^{-3}	1.21×10^{-07}	NR		
	10	*S10_1409344*	0.135	1.1×10^{-3}	1.01×10^{-07}	NR		
Flowering time	1	*S1_6693744*	0.142	1.82×10^{-4}	7.58×10^{-09}	NR		
	3	*S3_4495200*	0.169	4.04×10^{-5}	4.29×10^{-10}	NR		
Internode number	3	*S3_10820974*	0.157	4.09×10^{-3}	3.62×10^{-07}	NR		
	4	*S4_66398511*	0.105	2.59×10^{-3}	1.69×10^{-07}	NR		
	8	*S8_47566683*	0.138	2.59×10^{-3}	5.67×10^{-08}	NR		
	9	*S9_5579398*	0.098	5.9×10^{-3}	9.18×10^{-07}	NR		
Leaf angle	6	*S6_14800774*	0.148	1.42×10^{-7}	1.50×10^{-11}	NR		
		S6_15946838	0.159	1.39×10^{-7}	1.13×10^{-11}	NR		
		S6_27031083	0.143	1.39×10^{-7}	8.0×10^{-12}	NR		
		S6_28672419	0.167	5.05×10^{-8}	1.19×10^{-12}	NR		
		S6_34331229	0.156	1.39×10^{-7}	1.30×10^{-11}	NR		
Panicle length	2	*S2_12075851*	0.213	4.19×10^{-7}	3.09×10^{-11}	NR		
	5	*S5_53850782*	0.369	1.26×10^{-9}	1.55×10^{-14}	NR		
	7	*S7_61928412*	0.134	1.37×10^{-6}	1.78×10^{-10}	56624926	63754887	Hart et al., 2001
		S7_62484190	0.137	1.76×10^{-6}	3.03×10^{-10}			
Stem circumference	1	*S1_72437626*	0.120	3.5×10^{-6}	5.39×10^{-10}	NR		
		S1_72569239	0.136	1.98×10^{-6}	5.09×10^{-11}	NR		

Supplementary Table S5. Continued

Stem circumference	2	*S2_71792891*	0.127	2.01×10^{-6}	1.03×10^{-10}		NR	
	3	*S3_5779933*	0.194	1.52×10^{-6}	1.96×10^{-11}		NR	
	6	*S6_59370519*	0.117	3.1×10^{-6}	3.98×10^{-10}		NR	
	8	*S8_37735289*	0.113	2.55×10^{-6}	2.29×10^{-10}		NR	
Tiller Number	1	*S1_13349644*	0.158	4.67×10^{-5}	3.34×10^{-09}		NR	
		S1_13831837	0.110	4.88×10^{-5}	4.08×10^{-09}		NR	
	8	*S8_47476046*	0.127	4.56×10^{-5}	1.09×10^{-09}		NR	
Seed Number	1	*S1_50283671*	0.085	2.25×10^{-4}	3.39×10^{-08}		NR	
		S1_50747127	0.154	5.68×10^{-5}	4.01×10^{-09}		NR	
	2	*S2_70404546*	0.099	2.25×10^{-4}	3.39×10^{-08}		NR	
	3	*S3_58386533*	0.094	1.31×10^{-4}	1.1×10^{-08}		NR	
		S3_58594366	0.108	1.31×10^{-4}	1.39×10^{-08}		NR	
	4	*S4_6945677*	0.116	1.95×10^{-5}	9.16×10^{-10}		NR	
	6	*S6_54397150*	0.111	7.47×10^{-6}	1.04×10^{-10}	53666870	54514114	Nagaraja Reddy et al., 2013

Markers/regions overlapping with previously identified QTL (PIQTL) are highlighted in yellow.
Markers/regions outside of PIQTL regions but in LD with them are indicated in purple.
NR indicates newly identified SNP/region.

Supplemental Table S6. Association analysis results for Gibberellin (GA) candidate genes.

Trait	Chr	Marker	R^2	*P* value	*q* value	Gene	Marker Technology
SN	1	*S1_462670*	0.033	0.015	0.419	*Gibberellin 20-oxidase*	GBS
SN	1	*S1_584187*	0.033	0.038	0.500	*Gibberellin 3-oxidase*	GBS
SN	1	*S1_9420573*	0.003	0.626	0.802	*DELLA protein*	GBS
SN	1	*S1_12952253*	0.017	0.107	0.599	*Phytochrome-associated protein 4*	GBS
SN	1	*S1_26914874*	0.023	0.026	0.465	*Ent-copalyl diphosphate synthase*	SQNM
SN	2	*S2_420288*	0.006	0.255	0.683	*Gibberellin 2-oxidase*	GBS
SN	2	*S2_4376818*	0.018	0.035	0.494	*Gibberellin 20-oxidase*	GBS
SN	3	*S3_4286019*	0.021	0.041	0.507	*Gibberellin 3-oxidase*	GBS
SN	3	*S3_9288785*	0.002	0.470	0.758	*RING-box protein*	SQNM
SN	3	*S3_16765764*	0.020	0.106	0.599	*Gibberellin 2-oxidase*	GBS
SN	3	*S3_60650660*	0.011	0.287	0.692	*Gibberellic acid methyltransferase 1*	GBS
SN	3	*S3_69378415*	0.025	0.027	0.471	*Gibberellin 20-oxidase*	SQNM
SN	4	*S4_53752188*	0.036	0.006	0.352	*F-box protein GID2*	SQNM
SN	4	*S4_57923252*	0.054	0.004	0.323	*Cullin-1*	SQNM
SN	6	*S6_52092501*	0.017	0.086	0.579	*Gibberellin 2-oxidase*	GBS
SN	6	*S6_57049320*	0.087	5.95×10^{-7}	0.027	*Ent-kaurene synthase*	SQNM
SN	6	*S6_57056736*	0.046	1.24×10^{-4}	0.187	*Ent-kaurene synthase*	SQNM
SN	7	*S7_55486754*	0.015	0.087	0.580	*UFO-binding protein 1*	SQNM
SN	9	*S9_3387990*	0.000	0.969	0.858	*Phytochrome-associated protein 3*	GBS
SN	9	*S9_5489164*	0.002	0.488	0.764	*Gibberellin 2-oxidase*	SQNM
SN	9	*S9_53324845*	0.000	0.945	0.857	*Cytochrome P450 714D1*	GBS
SN	9	*S9_54884991*	0.008	0.332	0.712	*Gibberellin 2-oxidase*	GBS
SN	9	*S9_57266896*	0.003	0.639	0.804	*Gibberellin 2-oxidase*	GBS
SN	10	*S10_615886*	0.014	0.131	0.621	*Ent-kaurenoic acid oxidase*	GBS
SN	10	*S10_50453939*	0.008	0.408	0.739	*Ent-kaurene oxidase*	SQNM

Supplemental Table S6. Continued

Ht	1	*S1_462224*	0.001	0.598	0.717	*Gibberellin 20-oxidase*	GBS
Ht	1	*S1_577229*	0.011	0.166	0.470	*Gibberellin 3-oxidase*	GBS
Ht	1	*S1_9422612*	0.033	0.027	0.206	*DELLA protein*	GBS
Ht	1	*S1_12950371*	0.017	0.077	0.343	*Phytochrome-associated protein 4*	GBS
Ht	1	*S1_26916453*	0.002	0.787	0.768	*Ent-copalyl diphosphate synthase*	SQNM
Ht	2	*S2_420288*	0.001	0.598	0.717	*Gibberellin 2-oxidase*	GBS
Ht	2	*S2_4376643*	0.016	0.108	0.398	*Gibberellin 20-oxidase*	GBS
Ht	3	*S3_4290756*	0.043	0.001	0.029	*Gibberellin 3-oxidase*	GBS
Ht	3	*S3_9288955*	0.015	0.040	0.251	*RING-box protein*	SQNM
Ht	3	*S3_16765764*	0.028	0.055	0.292	*Gibberellin 2-oxidase*	GBS
Ht	3	*S3_60650660*	0.028	0.039	0.247	*Gibberellic acid methyltransferase 1*	GBS
Ht	3	*S3_69378263*	0.007	0.161	0.465	*Gibberellin 20-oxidase*	SQNM
Ht	4	*S4_53756499*	0.003	0.672	0.738	*F-box protein GID2*	SQNM
Ht	4	*S4_57923252*	0.005	0.642	0.730	*Cullin-1*	SQNM
Ht	6	*S6_52092827*	0.001	0.895	0.792	*Gibberellin 2-oxidase*	GBS
Ht	6	*S6_57048766*	0.008	0.128	0.426	*Ent-kaurene synthase*	SQNM
Ht	6	*S6_57056736*	0.010	0.092	0.373	*Ent-kaurene synthase*	SQNM
Ht	7	*S7_55484074*	0.017	0.025	0.199	*UFO-binding protein 1*	SQNM
Ht	9	*S9_3387212*	0.000	0.941	0.800	*Phytochrome-associated protein 3*	GBS
Ht	9	*S9_5482094*	0.002	0.514	0.689	*Gibberellin 2-oxidase*	SQNM
Ht	9	*S9_53326431*	0.025	0.058	0.299	*Cytochrome P450 714D1*	GBS
Ht	9	*S9_54884991*	0.058	0.001	0.022	*Gibberellin 2-oxidase*	GBS
Ht	9	*S9_57266896*	0.147	9.1×10^{-10}	0.000	*Gibberellin 2-oxidase*	GBS
Ht	10	*S10_615886*	0.005	0.495	0.682	*Ent-kaurenoic acid oxidase*	GBS
Ht	10	*S10_50449032*	0.013	0.143	0.445	*Ent-kaurene oxidase*	SQNM
TN	1	*S1_462209*	0.007	0.342	0.508	*Gibberellin 20-oxidase*	GBS
TN	1	*S1_584187*	0.039	0.040	0.273	*Gibberellin 3-oxidase*	GBS
TN	1	*S1_9420573*	0.009	0.334	0.505	*DELLA protein*	GBS

Supplemental Table S6. Continued

TN	1	*S1_12952253*	0.048	0.002	0.078	*Phytochrome-associated protein 4*	GBS
TN	1	*S1_26914874*	0.013	0.127	0.348	*Ent-copalyl diphosphate synthase*	SQNM
TN	2	*S2_420288*	0.001	0.745	0.637	*Gibberellin 2-oxidase*	GBS
TN	2	*S2_4376643*	0.008	0.331	0.503	*Gibberellin 20-oxidase*	GBS
TN	3	*S3_4286019*	0.045	0.002	0.074	*Gibberellin 3-oxidase*	GBS
TN	3	*S3_9288785*	0.004	0.254	0.456	*RING-box protein*	SQNM
TN	3	*S3_16765768*	0.015	0.227	0.438	*Gibberellin 2-oxidase*	GBS
TN	3	*S3_60650660*	0.036	0.017	0.209	*Gibberellic acid methyltransferase 1*	GBS
TN	3	*S3_69378415*	0.051	0.001	0.058	*Gibberellin 20-oxidase*	SQNM
TN	4	*S4_53752188*	0.011	0.134	0.354	*F-box protein GID2*	SQNM
TN	4	*S4_57923252*	0.012	0.277	0.471	*Cullin-1*	SQNM
TN	6	*S6_52092501*	0.001	0.851	0.662	*Gibberellin 2-oxidase*	GBS
TN	6	*S6_57049108*	0.001	0.525	0.577	*Ent-kaurene synthase*	SQNM
TN	6	*S6_57054908*	0.001	0.674	0.621	*Ent-kaurene synthase*	SQNM
TN	7	*S7_55486754*	0.022	0.035	0.264	*UFO-binding protein 1*	SQNM
TN	9	*S9_3387990*	0.019	0.066	0.304	*Phytochrome-associated protein 3*	GBS
TN	9	*S9_5489164*	0.008	0.111	0.331	*Gibberellin 2-oxidase*	SQNM
TN	9	*S9_53324845*	0.054	0.000	0.042	*Cytochrome P450 714D1*	GBS
TN	9	*S9_54884991*	0.029	0.022	0.226	*Gibberellin 2-oxidase*	GBS
TN	9	*S9_57266896*	0.013	0.133	0.354	*Gibberellin 2-oxidase*	GBS
TN	10	*S10_615886*	0.017	0.107	0.329	*Ent-kaurenoic acid oxidase*	GBS
TN	10	*S10_50453939*	0.028	0.062	0.301	*Ent-kaurene oxidase*	SQNM
SC	1	*S1_462241*	0.058	0.001	0.174	*Gibberellin 20-oxidase*	GBS
SC	1	*S1_577229*	0.001	0.885	0.726	*Gibberellin 3-oxidase*	GBS
SC	1	*S1_9422619*	0.018	0.096	0.427	*DELLA protein*	GBS
SC	1	*S1_12950758*	0.065	0.001	0.170	*Phytochrome-associated protein 4*	GBS
SC	1	*S1_26921352*	0.028	0.014	0.288	*Ent-copalyl diphosphate synthase*	SQNM

Supplemental Table S6. Continued

SC	2	*S2_420288*	0.002	0.598	0.667	*Gibberellin 2-oxidase*	GBS
SC	2	*S2_4376643*	0.004	0.466	0.626	*Gibberellin 20-oxidase*	GBS
SC	3	*S3_4290718*	0.048	0.000	0.163	*Gibberellin 3-oxidase*	GBS
SC	3	*S3_9288823*	0.034	0.014	0.286	*RING-box protein*	SQNM
SC	3	*S3_16765768*	0.012	0.261	0.533	*Gibberellin 2-oxidase*	GBS
SC	3	*S3_60650661*	0.042	0.011	0.274	*Gibberellic acid methyltransferase 1*	GBS
SC	3	*S3_69378415*	0.032	0.007	0.247	*Gibberellin 20-oxidase*	SQNM
SC	4	*S4_53752188*	0.035	0.007	0.249	*F-box protein GID2*	SQNM
SC	4	*S4_57923252*	0.033	0.036	0.348	*Cullin-1*	SQNM
SC	6	*S6_52092501*	0.008	0.324	0.564	*Gibberellin 2-oxidase*	GBS
SC	6	*S6_57048727*	0.021	0.032	0.341	*Ent-kaurene synthase*	SQNM
SC	6	*S6_57054656*	0.000	0.825	0.714	*Ent-kaurene synthase*	SQNM
SC	7	*S7_55486754*	0.034	0.005	0.228	*UFO-binding protein 1*	SQNM
SC	9	*S9_3387412*	0.036	0.004	0.206	*Phytochrome-associated protein 3*	GBS
SC	9	*S9_5485728*	0.015	0.188	0.494	*Gibberellin 2-oxidase*	SQNM
SC	9	*S9_53332208*	0.021	0.032	0.340	*Cytochrome P450 714D1*	GBS
SC	9	*S9_54884991*	0.006	0.455	0.622	*Gibberellin 2-oxidase*	GBS
SC	9	*S9_57261485*	0.003	0.344	0.574	*Gibberellin 2-oxidase*	GBS
SC	10	*S10_615935*	0.033	0.008	0.251	*Ent-kaurenoic acid oxidase*	GBS
SC	10	*S10_50454613*	0.022	0.024	0.322	*Ent-kaurene oxidase*	SQNM
PE	1	*S1_462224*	0.005	0.311	0.515	*Gibberellin 20-oxidase*	GBS
PE	1	*S1_577229*	0.009	0.277	0.500	*Gibberellin 3-oxidase*	GBS
PE	1	*S1_9422612*	0.019	0.109	0.382	*DELLA protein*	GBS
PE	1	*S1_12950371*	0.028	0.018	0.210	*Phytochrome-associated protein 4*	GBS
PE	1	*S1_26916453*	0.026	0.025	0.231	*Ent-copalyl diphosphate synthase*	SQNM
PE	2	*S2_420288*	0.006	0.300	0.510	*Gibberellin 2-oxidase*	GBS
PE	2	*S2_4376818*	0.015	0.063	0.320	*Gibberellin 20-oxidase*	GBS

Supplemental Table S6. Continued

PE	3	*S3_4290756*	0.022	0.036	0.262	*Gibberellin 3-oxidase*	GBS
PE	3	*S3_9288955*	0.018	0.020	0.218	*RING-box protein*	SQNM
PE	3	*S3_16765764*	0.009	0.431	0.556	*Gibberellin 2-oxidase*	GBS
PE	3	*S3_60650660*	0.034	0.030	0.248	*Gibberellic acid methyltransferase 1*	GBS
PE	3	*S3_69378263*	0.017	0.028	0.243	*Gibberellin 20-oxidase*	SQNM
PE	4	*S4_53756499*	0.017	0.118	0.391	*F-box protein GID2*	SQNM
PE	4	*S4_57923252*	0.001	0.893	0.674	*Cullin-1*	SQNM
PE	6	*S6_52092827*	0.007	0.393	0.544	*Gibberellin 2-oxidase*	GBS
PE	6	*S6_57048766*	0.011	0.076	0.340	*Ent-kaurene synthase*	SQNM
PE	6	*S6_57054908*	0.012	0.054	0.303	*Ent-kaurene synthase*	SQNM
PE	7	*S7_55484074*	0.079	0.000	0.044	*UFO-binding protein 1*	SQNM
PE	9	*S9_3387212*	0.049	0.004	0.136	*Phytochrome-associated protein 3*	GBS
PE	9	*S9_5482094*	0.005	0.351	0.530	*Gibberellin 2-oxidase*	SQNM
PE	9	*S9_53326431*	0.075	0.000	0.076	*Cytochrome P450 714D1*	GBS
PE	9	*S9_54884991*	0.026	0.033	0.257	*Gibberellin 2-oxidase*	GBS
PE	9	*S9_57261485*	0.013	0.067	0.326	*Gibberellin 2-oxidase*	GBS
PE	10	*S10_615886*	0.043	0.005	0.142	*Ent-kaurenoic acid oxidase*	GBS
PE	10	*S10_50449032*	0.011	0.184	0.446	*Ent-kaurene oxidase*	SQNM
FT	1	*S1_462102*	0.008	0.493	0.711	*Gibberellin 20-oxidase*	GBS
FT	1	*S1_584187*	0.012	0.424	0.686	*Gibberellin 3-oxidase*	GBS
FT	1	*S1_9422612*	0.019	0.158	0.603	*DELLA protein*	GBS
FT	1	*S1_12950758*	0.073	0.003	0.215	*Phytochrome-associated protein 4*	GBS
FT	1	*S1_26914874*	0.038	0.008	0.320	*Ent-copalyl diphosphate synthase*	SQNM
FT	2	*S2_420288*	0.011	0.222	0.630	*Gibberellin 2-oxidase*	GBS
FT	2	*S2_4377486*	0.021	0.218	0.628	*Gibberellin 20-oxidase*	GBS
FT	3	*S3_4286019*	0.061	0.001	0.204	*Gibberellin 3-oxidase*	GBS
FT	3	*S3_9288955*	0.024	0.019	0.395	*RING-box protein*	SQNM

Supplemental Table S6. Continued

FT	3	*S3_16765768*	0.025	0.060	0.497	*Gibberellin 2-oxidase*	GBS
FT	3	*S3_60650660*	0.037	0.055	0.489	*Gibberellic acid methyltransferase 1*	GBS
FT	3	*S3_69377305*	0.007	0.428	0.688	*Gibberellin 20-oxidase*	SQNM
FT	4	*S4_53752145*	0.013	0.358	0.658	*F-box protein GID2*	SQNM
FT	4	*S4_57923252*	0.052	0.022	0.407	*Cullin-1*	SQNM
FT	6	*S6_52092501*	0.030	0.047	0.476	*Gibberellin 2-oxidase*	GBS
FT	6	*S6_57044261*	0.035	0.010	0.341	*Ent-kaurene synthase*	SQNM
FT	6	*S6_57054656*	0.007	0.209	0.625	*Ent-kaurene synthase*	SQNM
FT	7	*S7_55486754*	0.055	0.002	0.207	*UFO-binding protein 1*	SQNM
FT	9	*S9_3387990*	0.047	0.007	0.305	*Phytochrome-associated protein 3*	GBS
FT	9	*S9_5485729*	0.058	0.014	0.365	*Gibberellin 2-oxidase*	SQNM
FT	9	*S9_53324845*	0.016	0.140	0.589	*Cytochrome P450 714D1*	GBS
FT	9	*S9_54885754*	0.023	0.125	0.578	*Gibberellin 2-oxidase*	GBS
FT	9	*S9_57260889*	0.024	0.058	0.494	*Gibberellin 2-oxidase*	GBS
FT	10	*S10_615858*	0.013	0.122	0.576	*Ent-kaurenoic acid oxidase*	GBS
FT	10	*S10_50454613*	0.017	0.103	0.556	*Ent-kaurene oxidase*	SQNM
IN	1	*S1_462670*	0.033	0.031	0.528	*Gibberellin 20-oxidase*	GBS
IN	1	*S1_584187*	0.019	0.242	0.685	*Gibberellin 3-oxidase*	GBS
IN	1	*S1_9422612*	0.010	0.354	0.722	*DELLA protein*	GBS
IN	1	*S1_12952253*	0.032	0.040	0.540	*Phytochrome-associated protein 4*	GBS
IN	1	*S1_26914874*	0.018	0.094	0.611	*Ent-copalyl diphosphate synthase*	SQNM
IN	2	*S2_420288*	0.003	0.502	0.759	*Gibberellin 2-oxidase*	GBS
IN	2	*S2_4376643*	0.024	0.150	0.650	*Gibberellin 20-oxidase*	GBS
IN	3	*S3_4286019*	0.033	0.020	0.496	*Gibberellin 3-oxidase*	GBS
IN	3	*S3_9292355*	0.023	0.122	0.635	*RING-box protein*	SQNM
IN	3	*S3_16765768*	0.007	0.329	0.714	*Gibberellin 2-oxidase*	GBS
IN	3	*S3_60650660*	0.031	0.107	0.625	*Gibberellic acid methyltransferase 1*	GBS

Supplemental Table S6. Continued

IN	3	*S3_69377305*	0.006	0.438	0.745	*Gibberellin 20-oxidase*	SQNM
IN	4	*S4_53756922*	0.006	0.584	0.772	*F-box protein GID2*	SQNM
IN	4	*S4_57923252*	0.075	0.004	0.350	*Cullin-1*	SQNM
IN	6	*S6_52092501*	0.051	0.004	0.348	*Gibberellin 2-oxidase*	GBS
IN	6	*S6_57044227*	0.024	0.040	0.542	*Ent-kaurene synthase*	SQNM
IN	6	*S6_57056736*	0.002	0.533	0.766	*Ent-kaurene synthase*	SQNM
IN	7	*S7_55486754*	0.049	0.004	0.348	*UFO-binding protein 1*	SQNM
IN	9	*S9_3387990*	0.009	0.219	0.677	*Phytochrome-associated protein 3*	GBS
IN	9	*S9_5488006*	0.012	0.280	0.696	*Gibberellin 2-oxidase*	SQNM
IN	9	*S9_53332879*	0.011	0.396	0.733	*Cytochrome P450 714D1*	GBS
IN	9	*S9_54884991*	0.023	0.090	0.606	*Gibberellin 2-oxidase*	GBS
IN	9	*S9_57260889*	0.011	0.256	0.688	*Gibberellin 2-oxidase*	GBS
IN	10	*S10_615826*	0.015	0.113	0.629	*Ent-kaurenoic acid oxidase*	GBS
IN	10	*S10_50452521*	0.020	0.112	0.628	*Ent-kaurene oxidase*	SQNM
PL	1	*S1_462483*	0.005	0.621	0.559	*Gibberellin 20-oxidase*	GBS
PL	1	*S1_577229*	0.021	0.045	0.214	*Gibberellin 3-oxidase*	GBS
PL	1	*S1_9422612*	0.029	0.025	0.185	*DELLA protein*	GBS
PL	1	*S1_12952253*	0.053	0.001	0.063	*Phytochrome-associated protein 4*	GBS
PL	1	*S1_26927740*	0.030	0.015	0.160	*Ent-copalyl diphosphate synthase*	SQNM
PL	2	*S2_420288*	0.000	0.993	0.660	*Gibberellin 2-oxidase*	GBS
PL	2	*S2_4377486*	0.078	0.001	0.049	*Gibberellin 20-oxidase*	GBS
PL	3	*S3_4290718*	0.020	0.044	0.212	*Gibberellin 3-oxidase*	GBS
PL	3	*S3_9292303*	0.018	0.118	0.280	*RING-box protein*	SQNM
PL	3	*S3_16765764*	0.012	0.279	0.405	*Gibberellin 2-oxidase*	GBS
PL	3	*S3_60651796*	0.025	0.042	0.210	*Gibberellic acid methyltransferase 1*	GBS
PL	3	*S3_69377305*	0.010	0.221	0.367	*Gibberellin 20-oxidase*	SQNM
PL	4	*S4_53756922*	0.025	0.035	0.201	*F-box protein GID2*	SQNM

Supplemental Table S6. Continued

PL	4	*S4_57923252*	0.009	0.459	0.497	*Cullin-1*	SQNM
PL	6	*S6_52092827*	0.031	0.012	0.148	*Gibberellin 2-oxidase*	GBS
PL	6	*S6_57048727*	0.035	0.006	0.121	*Ent-kaurene synthase*	SQNM
PL	6	*S6_57056736*	0.013	0.049	0.219	*Ent-kaurene synthase*	SQNM
PL	7	*S7_55484074*	0.050	0.000	0.044	*UFO-binding protein 1*	SQNM
PL	9	*S9_3387869*	0.081	0.001	0.062	*Phytochrome-associated protein 3*	GBS
PL	9	*S9_5485728*	0.027	0.077	0.240	*Gibberellin 2-oxidase*	SQNM
PL	9	*S9_53332220*	0.027	0.024	0.183	*Cytochrome P450 714D1*	GBS
PL	9	*S9_54884988*	0.026	0.033	0.197	*Gibberellin 2-oxidase*	GBS
PL	9	*S9_57266896*	0.003	0.625	0.560	*Gibberellin 2-oxidase*	GBS
PL	10	*S10_615971*	0.049	0.001	0.064	*Ent-kaurenoic acid oxidase*	GBS
PL	10	*S10_50452521*	0.023	0.031	0.194	*Ent-kaurene oxidase*	SQNM
LA	1	*S1_462670*	0.029	0.034	0.204	*Gibberellin 20-oxidase*	GBS
LA	1	*S1_584187*	0.025	0.087	0.294	*Gibberellin 3-oxidase*	GBS
LA	1	*S1_9422612*	0.029	0.037	0.209	*DELLA protein*	GBS
LA	1	*S1_12950975*	0.031	0.015	0.151	*Phytochrome-associated protein 4*	GBS
LA	1	*S1_26914874*	0.034	0.005	0.092	*Ent-copalyl diphosphate synthase*	SQNM
LA	2	*S2_420288*	0.001	0.690	0.606	*Gibberellin 2-oxidase*	GBS
LA	2	*S2_4376818*	0.015	0.066	0.265	*Gibberellin 20-oxidase*	GBS
LA	3	*S3_4287451*	0.031	0.008	0.115	*Gibberellin 3-oxidase*	GBS
LA	3	*S3_9288785*	0.010	0.081	0.287	*RING-box protein*	SQNM
LA	3	*S3_16765764*	0.015	0.222	0.394	*Gibberellin 2-oxidase*	GBS
LA	3	*S3_60650660*	0.060	0.002	0.065	*Gibberellic acid methyltransferase 1*	GBS
LA	3	*S3_69378415*	0.010	0.243	0.408	*Gibberellin 20-oxidase*	SQNM
LA	4	*S4_53752188*	0.017	0.088	0.296	*F-box protein GID2*	SQNM
LA	4	*S4_57923252*	0.008	0.479	0.532	*Cullin-1*	SQNM
LA	6	*S6_52092501*	0.004	0.574	0.569	*Gibberellin 2-oxidase*	GBS

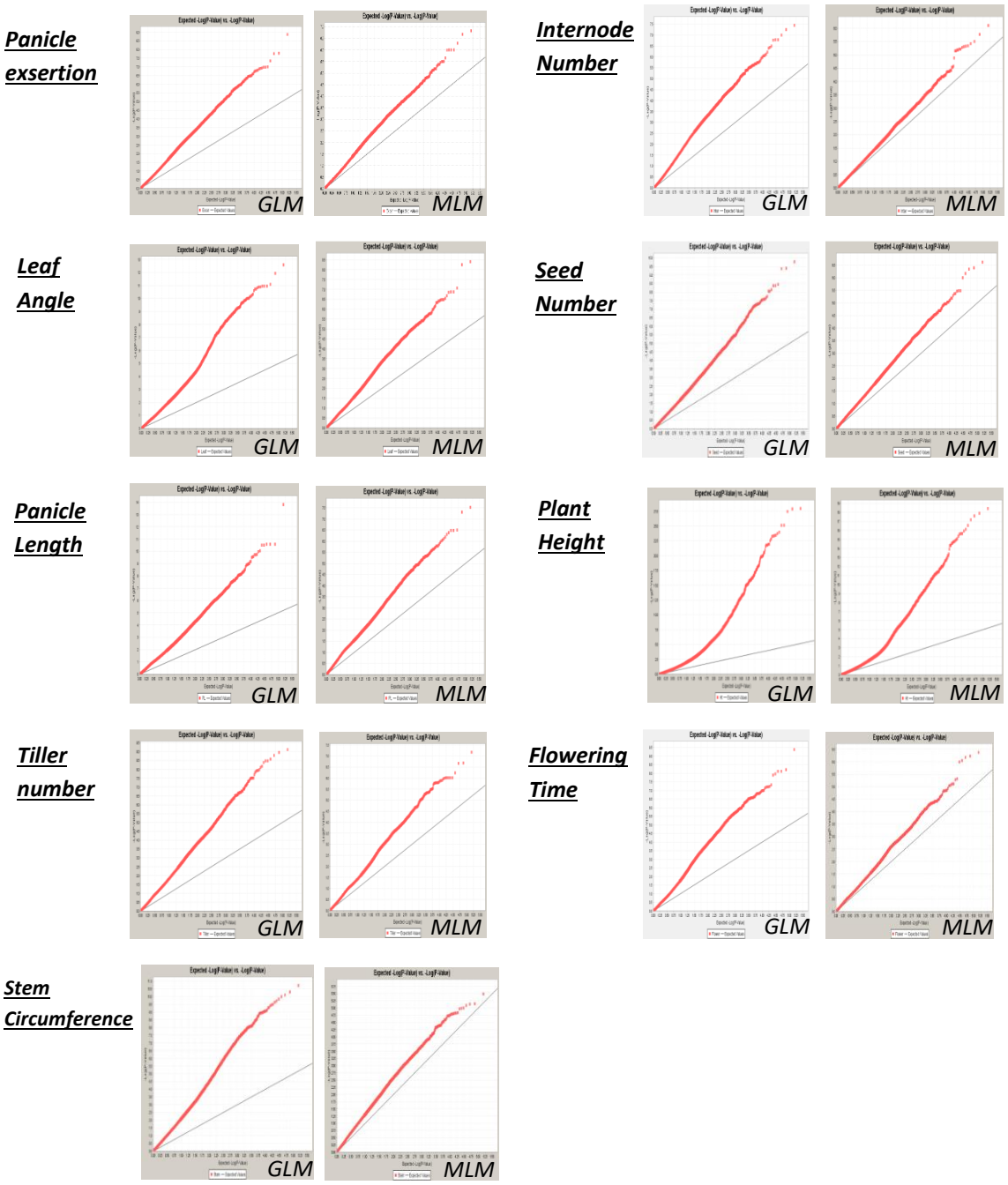
Supplemental Table S6. Continued

LA	6	*S6_57041349*	0.026	0.021	0.169	*Ent-kaurene synthase*	SQNM
LA	6	*S6_57054908*	0.001	0.509	0.544	*Ent-kaurene synthase*	SQNM
LA	7	*S7_55486754*	0.009	0.272	0.427	*UFO-binding protein 1*	SQNM
LA	9	*S9_3386381*	0.048	0.008	0.115	*Phytochrome-associated protein 3*	GBS
LA	9	*S9_5484119*	0.038	0.018	0.158	*Gibberellin 2-oxidase*	SQNM
LA	9	*S9_53324845*	0.032	0.011	0.132	*Cytochrome P450 714D1*	GBS
LA	9	*S9_54884991*	0.028	0.034	0.203	*Gibberellin 2-oxidase*	GBS
LA	9	*S9_57266896*	0.052	0.001	0.037	*Gibberellin 2-oxidase*	GBS
LA	10	*S10_615886*	0.030	0.021	0.168	*Ent-kaurenoic acid oxidase*	GBS
LA	10	*S10_50454613*	0.023	0.029	0.191	*Ent-kaurene oxidase*	SQNM

PE, panicle exsertion; FT, flowering time; LA, leaf angle; SC, stem circumference; PH, plant height; SN, seed number; IN, internode number; TN, tiller number; PL, panicle length; One marker with lowest *P*&*q* value within each trait association analysis was listed.

Supplemental Table S7. Candidate genes in the refined interval on chromosome 7 associated with variation in leaf angle

ID	Chr	Gene Begin	Gene End	Description	Reference
Sb07g023360	7	58294657	58295592	ZF-HD homeobox protein, putative, Response to abiotic stress， Response to ABA	Xu et al. 2014
Sb07g023380	7	58315159	58317795	CAMK_CAMK_like.38 - CAMK includes calcium/calmodulin depedent protein kinases, expressed	Yang and Komatsu 2000
Sb07g023575	7	58462919	58464082	AP2 domain containing protein, expressed	Jiang et al. 2012
Sb07g023803	7	58766072	58766443	AP2 domain containing protein, expressed	
Sb07g024110	7	59108697	59114721	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyl transferase SPINDLY, putative, expressed	Shimada et al. 2006
Sb07g024740	7	59743463	59744100	OsSAUR36 - Auxin-responsive SAUR gene family member	Kant et al. 2009
Sb07g024750	7	59746707	59747144	OsSAUR36 - Auxin-responsive SAUR gene family member	



Supplemental Figure S1. Comparison of QQ plots for general linear model (GLM) and mixed linear model (MLM) results for each trait.

CHAPTER 3. DISCOVERY OF A HORMONAL GENE THAT CONTROLS SEED NUMBER PER INFLORESCENCE, AN IMPORTANT DETERMINANT OF GRAIN YIELD

Abstract

Seed number per inflorescence, one of the major components of grain yield, has been extensively explored in cereal crops. In sorghum, the fifth most widely grown cereal species in the world, up to 70% of the observed grain yield can be explained by variation in seed number per inflorescence. However, there are important knowledge gaps about the genetic mechanisms that control this economically important trait. Recently, GWAS has been verified as a powerful method to identify the causal genes underlying variation in seed size in rice and drought tolerance in maize seedlings, and the effects of these genes were successfully validated through transformation studies. Following a similar approach, *SbKS3*, associated with variation in seed number per panicle and discovered by GWAS, was validated by multiple experiments that included differential expression in multiple tissues, comparative genomics and transgenic studies. The two haplotypes of *SbKS3* were tested by overexpression analysis in transgenic rice, to verify its function in controlling seed number. Haplotype I is the functional allele that increased seed number by altering panicle branch length. This gene and the new knowledge generated herein could be utilized by genetic engineering, editing approaches or in breeding programs, for yield enhancement of sorghum and other important grass species such as rice, maize and wheat.

Introduction

In cereal crops, grain yield per plant is determined by three components: number of inflorescences per plant, seed number per inflorescence, and grain weight (Egli 2017). Among them, seed number per panicle (also known as spikelet number), is the most effective determinant of yield across cereal species (Gerik et al. 2003; Peltonen-Sainio et al. 2007; Saeed et al. 1986; van Oosterom and Hammer 2008).

The quantitative genetic basis of seed number has been well demonstrated using natural allelic variation, and genes controlling this trait have been successfully cloned in cereal crops (Ashikari et al. 2005; Bartrina et al. 2011; Gu et al. 2015; Sreenivasulu and Schnurbusch 2012). Based on this knowledge, cytokinins were identified as important determinants of seed number per inflorescence, since higher levels of this hormonal group are associated with greater seed number and yield (Ashikari et al. 2005; Bartrina et al. 2011; Gu et al. 2015). E.g., *Gn1a* is a cytokinin oxidase/dehydrogenase (*OsCKX2*) gene that degrades the active hormone, resulting in lower seed number and reduced overall yield (Ashikari et al. 2005).Other hormonal groups, such as gibberellins, also play a significant role in regulating seed number due to their crosstalk with other phytohormones. For instance, the gibberellin biosynthesis gene Grain Number per Panicle1 (*GNP1)* regulates seed number through a KNOX protein-mediated crosstalk between gibberellins and cytokinins (Wu et al. 2016).

Genes that control seed number have been frequently reported to have pleiotropic effects on other important agronomic traits such as tiller number, days to heading, and plant height. E.g., the *WEALTHY FARMER'S PANICLE* (*IPA1/WFP*) gene, identified by map-based cloning, not only affected seed number but also had a pleiotropic effect on tiller number. This gene is involved in shoot branching determination during the vegetative stage, and then affects primary

branch number in young panicles (Miura et al. 2010). Days to heading and plant height are other common pleiotropic traits affected by seed number genes. *Ghd7,* encoding a CCT domain protein, has major effects on an array of traits in rice, including number of grains per panicle, plant height and heading date. Enhanced expression of *Ghd7* under long-day conditions delays heading and increases plant height and panicle size (Xue et al. 2008). *DTH8,* encoding a putative HAP3 subunit of the CCAAT-box-binding transcription factor, significantly increased days to heading, plant height, and number of grains per panicle in CSSL61 (a chromosome segment substitution line that carries the non-functional *DTH8* allele) under long-day conditions (Wei et al. 2010).

In sorghum, up to 70% of the observed grain yield can be explained by variation in seed number per panicle (Gerik et al., 2003). However, significant knowledge gaps remain about the genetic mechanisms that determine seed number per inflorescence in this species. Studies attempting to detect genomic regions associated with sorghum seed number have been limited (Rami et al. 1998; Nagaraja Reddy et al. 2013; Boyles et al. 2016), and no gene underlying these loci have been identified. Zhao et al. (2016) conducted a GWAS on seed number, in which the five most significant associations corresponded to SNPs localized within a single hormone gene (*Sb06g028210).* Considering this preliminary evidence about the potential role of *Sb06g028210* on seed number determination, a series of experiments were conducted to accomplish the following objectives: (i) to confirm the association of *SbKS3* with seed number variation; (ii) to determine the biochemical activity of *Sb06g028210*; (iii) to validate its function in seed number determination; and (iv) to discover the expression profile of the two alternative *SbKS3* haplotypes.

Materials and Methods

SbKS3-based association analysis with seed number variation

Primers to amplify *SbKS3* were designed using Prime 3.0 (https://www.ncbi.nlm.nih.gov/tools/primer-blast), and the polymorphisms discovered from a subset of 315 lines were subsequently genotyped in the entire SAP by sequencing amplicons using Illumina ABI3730. The sequences were assembled and aligned to the reference genome using SEQUENCHER version 5.4.6 (Gene Codes Corporation, Ann, Arbor, MI, USA). Polymorphisms, including SNPs and indels, were identified and only those markers with missing data <20% and minor allele frequency >5% were used in the *SbKS3*-based association analysis performed according to Zhao et al. (2016). Pairwise LD within *SbKS3* was calculated and plotted using the R package of LDheatmap (Shin et al. 2006). Two haplotypes, corresponding to high and low seed number groups, were discovered using Flapjack-1.15.04.27 (Milne et al. 2010).

Recombinant biochemical analysis

The biochemical characterization of *SbKS1* and *SbKS3* was carried out through a modular metabolic engineering system (Cyr et al. 2007). Briefly, *SbKS1* or *SbKS3* was expressed as a pseudo-mature enzyme in *Escherichia coli (E. coli)* engineered to produce *ent-CPP*, resulting in the formation of *ent-kaurene* to demonstrate KS activity. The coding sequences of *SbKS1* and *SbKS3* were identified using the reference sorghum genome (www.phytozome.net) and artificially synthesized in Thermo Fisher Scientific (former *life*) company. The genes were amplified by PCR for directional topoisomerization into the pENTR/SD/D-TOPO vector (Invitrogen). This construct was verified by sequencing and transferred via directional recombination into the pDEST17 expression vector (Invitrogen). The expression constructs were transformed into C41 OverExpress *E.coli* cells (Lucigen) harboring a pGG-DEST construct with

both a *GGPP synthase* and *ent-CPP synthase*, which was created by recombination into the DESTination cassette by the Gateway cloning system (Invitrogen). To increase metabolic flux toward the isoprenoid pathway, pIRS was co-transformed into the C41 cells as well (Morrone et al. 2010). The resulting recombinant strains were cultured in 50 mL of liquid Terrific Broth medium at 37°C until they reached an optical density at 600 nm (OD600 nm) of *c*. 0.6. The cultures were then shifted to 16°C for 30 minutes before induction with IPTG (isopropyl-β-d-thiogalactopyranoside) at 0.5 mM and addition of phosphate buffer (pH 7.0) to 100 mM, pyruvate to 50 mM, and MgCl2 to 1 mM. Fermentation was continued at 16°C for another 72 h. The cultures were then extracted with an equal volume of hexane, which was dried under nitrogen, and the endogenously dephosphorylated products were re-suspended in 1 mL of hexane. This concentrated extract was analyzed by gas chromatography (GC) carried out on a Varian (Palo Alto, CA, USA) 3900 GC with Saturn 2100 ion trap mass spectrometer (MS) in electron ionization (70 eV) mode. Samples (1 μl) were injected in splitless mode at 50°C and, after holding for 3 min at 50°C, the oven temperature was raised at a rate of 15°C min^{-1} to 300°C, where it was held for an additional 3 min. MS data from 90 to 600 *m/z* were collected starting 12 min after injection until the end of the run. The production of *ent*-kaurene was verified by comparison of mass spectra and retention time with an authentic standard (enzymatically produced by the characterized *CPS* and *KS* from *Arabidopsis thaliana*).

Function validation from two independent populations

Two F2 populations were created by crossing sorghum lines (PI533955 x PI552861 and PI629059 x PI656013) contrasting in *SbKS3* haplotypes. In 2015, approximately 1,500 seeds of each F2 population were planted at the ISU Agricultural Engineering and Agronomy farm (Ames, IA) on June 1st, 2015. Approximately 20 days after planting, leaf tissue was collected

from every plant for DNA extraction as described in Zhao et al. (2016). After physiological maturity, 250 individuals per population were randomly selected, manually harvested, barcoded and phenotyped for panicle length. Each panicle was subsequently threshed and cleaned by hand to reduce seed losses associated with mechanical threshing and air blowing procedures. A mechanical seed counter was utilized to obtain the number of seeds per panicle. DNA from the phenotyped plants was amplified with *SbKS3* gene-specific markers (Supplemental Table S1) and those amplicons were sequenced using Illumina ABI3730. Sequence alignment and polymorphism detection were conducted using SEQUENCHER version 5.4.6 (Gene Codes Corporation, Ann, Arbor, MI, USA). Single marker association analysis was performed as implemented in Tassel 5.0, including panicle length as a covariate (Bradbury et al. 2007; Zhang et al. 2010).

Rice transformation and phenotypic analysis

The cDNA of *SbKS3* was synthesized for haplotype I (2586 bp) and haplotype II (2553 bp), and independently cloned into the *Bam*HI and *Sac*I sites of the pCambia1300 vector under the control of the maize Ubiquitin promoter (GenScript, Piscataway, NJ). These plasmids carrying alternative alleles of SbKS3 were transformed into *Agrobacterium tumefaciens* strain EHA105 and each of the two pUbi: *SbKS3* expression cassettes were subsequently introduced into rice, Kitaake variety, at the ISU Plant Transformation Facility.

Transgenic T_0, T_1, and T_2 rice plants were grown under a 12-h-light/12-h-dark cycle, and molecularly characterized using *SbKS3* specific primers. T_0 and T_1 transgenic plants were self-pollinated and seedlings of T_1 with a single *SbKS3* insertion (identified by progeny test with a null segregant: transgenic line ratio of 1:3) were subsequently identified to advance to T_2 generation. T_2 homozygous transgenic lines for three independent events of each *SbKS3*

haplotype were selected by a progeny test. Null segregants and the selected transgenic lines for each event were then transplanted side by side into a cultivation box filled with flooded top soil.

Phenotypic data collection and analysis for the T_2 generation were performed as follows. Flowering time was recorded as the number of days from planting to panicle heading at a size of approx. 2cm. All other traits were collected at harvest time. Plant height was measured as the distance from the ground to the panicle notch. Exsertion was the length from the flag leaf ligule to the panicle notch. Panicle length was determined from the panicle notch to the tip of panicle. Seed related traits were characterized after drying at 37 ℃ for 72 hours, using only the main panicle. Primary branch number, branch length, seed number per branch, and aborted seed number per branch were manually collected. Total seed weight per inflorescence refers to the weight of all seeds after threshing, and thus, it does not include the weight attributed to panicle branches. Panicle branch length was the sum of each primary branch length. Biomass was measured as the above-ground weight for the entire plant after drying for 7 days at 72℃. A general linear model (PROC GLM) in SAS version 9.2 (SAS Institute, 2008) was applied using pooled data from the three independent events for each haplotype to compare: a) the null segregants and transgenic lines within each haplotype, and b) the transgenic lines between haplotypes.

SbKS3 expression profile in sorghum

The expression profile of *SbKS3* was characterized in two sets of experiments conducted under greenhouse and field conditions. In the greenhouse, two sorghum lines (PI533955 and PI552861), contrasting in seed number and haplotypes, were utilized to obtain RNA samples from multiple tissues (leaf, stem, root) under different developmental stages (three-leaf stage, transition from vegetative to reproductive growth at approx. seven-leaf stage, and post-transition

with an apical meristem of approx. 4 cm). Additionally, meristems were collected at five time points between the vegetative-to-reproductive transition and panicle heading, when the meristem length was approx. 2 mm, 1 cm, 5 cm, 12 cm and 22 cm. In the field, an expanded set of ten lines from the sorghum association panel contrasting in seed number and *SbKS3* haplotypes were planted and sampled at the three-leaf stage, on the vegetative-to-reproductive transition, and on the post-transition (panicle approx. 4cm) period.

For both sets of samples, reverse transcript-PCR (RT-PCR) was used to characterize the *SbKS3* expression profile. Total RNA was extracted using RNeasy Plant Mini Kit, followed by DNase treatment (TURBO DNA-free Kit). RNA quality and quantity were determined using Nanodrop and the total RNA (0.5ug) was reverse transcribed into cDNA [SuperScript III First-Strand Synthesis System (Invitrogen)]. RT-PCR was performed using a *SbKS3* primer (Supplemental Table S1), and ubiquitin as a control.

In silico analysis

Rice and maize orthologs of *Sb06g028210* and *Sb06g028220* were identified *in silico* using the sorghum protein sequences obtained from phytozome V2.1 (Paterson et al. 2009), and blasting them against the National Center for Biotechnology Information (NCBI) databases using BLASTP. The multiple protein sequence alignment and phylogenetic tree were generated using Clustal Omega (Sievers et al. 2011). The collinear relationship between sorghum, rice and maize was established based on information obtained from http://chibba.agtec.uga.edu/duplication (Zhang et al. 2015). The collinear relationship with wheat and barley group 2 was based on mapped ESTs stored in the Graingene database (www.wheat.pw.usda.gov).

Results

SbKS3 gene-specific association analysis with seed number variation

SbKS3 was initially identified as a candidate gene controlling seed number by a GWAS (Zhao et al. 2016), in which five SNPs within *SbKS3* were significantly associated with the trait. The most significant of the five SNPs explained 9% of the phenotypic variation. To validate this association, we sequenced the entire *SbKS3* gene across the sorghum association panel (over 300 accessions). Sixty-four polymorphisms, including Indels and SNPs, were originally detected and 42 of those (missing data < 20%, and minor allele frequency > 0.05) were actually utilized for the *SbKS3*-specific association study. Of the 42 markers, 24 were located in introns, one in the 3' UTR, and 17 in exons one, four, seven, eight, nine, and ten (Fig. 1a, b). As shown in Fig. 1c, all polymorphisms within *SbKS3* were in strong linkage disequilibrium (LD) ($r^2 > 0.8$) except for two SNPs ($r^2 < 0.2$). This explains the poor association detected between these two SNPs and the trait, and thus were excluded from further studies.

The *SbKS3*-specific association analysis demonstrated that several polymorphisms exhibit stronger associations than the ones detected in the original GWAS (Zhao et al. 2016), confirming the correlation between allelic variants in this gene and seed number (Fig. 2a). Due to the high LD, two haplotypes were clearly identified that corresponded to the high and low seed number groups (Fig. 1d). As shown in Fig. 1e, lines with Haplotype I had a significantly higher seed number than those with Haplotype II ($P < 0.0001$; R^2=11%).

Sb06g028210 functions as an ent-kaurene synthase (KS)

Ent-kaurene synthase (KS) functions in plastids (Yamaguchi 2008), where it converts ent-Copalyl diphosphate to ent-kaurene as part of the first few steps in the gibberellin biosynthesis pathway (Fig. 2a). Considering that the *Sb06g028210* protein is significantly shorter

than the ones encoded by the other two KS genes (*Sb06g028220, Sb05g022320*) (Supplemental Fig. S1), we conducted *in vitro* experiments to express *Sb06g028210* haplotype I and *Sb06g028220* separately in *E. coli* and investigate their enzymatic activity. *Ent*-kaurene was the chemical compound synthesized *in vitro* (Fig. 2a, 2b, 2c), which demonstrates that both *Sb06g028210 and Sb06g028220* encodes a functional Ent-kaurene synthase with a catalytic activity to convert *Ent*-Copalyl diphosphate (substrate) into Ent-kaurene (product). Considering the nomenclature used by Ordonio et al. (2015) in which "*SbKS1*" (*Sb06g028220*) and "*SbKS2*" (*Sb05g022320*) were identified, *Sb06g028210* is here designated as "*SbKS3*".

Function validation of *SbKS3* with seed number by independent populations.

SbKS3's effect on seed number variation was initially validated by testing the genotypic and phenotypic correlation using two independent F2 populations with parents contrasting in *SbKS3* haplotypes. Single marker analysis results confirmed the correlation between seed number variation and SNPs within *SbKS3* when panicle length was used as a covariate in the analysis (Supplemental Table S2).

Overexpression of two alternative alleles of *SbKS3* in rice

Genetic transformation was used to further validate the role of *SbKS3* in seed number determination and compare the two alternative alleles. Considering the technical challenges of sorghum transformation and the similarity between rice and sorghum inflorescences, rice transformation was implemented in this objective to overexpress the two alternative haplotypes of *SbKS3* driven by the Ubiquitin promoter. Considering that introns were not included in the

transformation construct, the number of polymorphisms under investigation was reduced and limited to exon-specific SNPs (Supplemental Fig. S2).

Seed number related traits were collected in the T_2 generation and compared within and between haplotypes, pooling data from three independent events per haplotype. Transgenic lines carrying the haplotype I allele exhibited greater seed number than the corresponding null segregants ($P < 0.001$), while there was no significant difference between transgenic lines and null segregants carrying haplotype II ($P = 0.18$) (Fig. 3a and Table 1). Considering that there was no significant difference between null segregants generated from haplotype I and II events ($P = 0.75$), we can conclude that the two transformation processes were comparable. Additionally, haplotype I transgenic lines showed greater seed number than those carrying haplotype II ($P = 0.0014$, $R^2 = 8.9\%$), confirming the GWAS and *SbKS3*-specific association analysis.

Seed weight per inflorescence followed the same trend observed in the seed number results (Fig. 3b; Table 1); i.e., there was no significant difference between null segregants and transgenic lines within haplotype II ($P = 0.57$) or between null segregants originated from events with either haplotype ($P = 0.42$). However, significant differences were detected between null segregants and transgenic lines carrying the haplotype I allele ($P < 0.0001$) and between transgenic lines with haplotype I and II ($P < 0.0001$). These results illustrate the contribution of seed number to the final weight per inflorescence.

To better understand how morphological changes affected seed number, the panicle was dissected into its components: primary branch number, primary branch length, seed number per primary branch, panicle branch length, and seed number per unit of branch length. This analysis demonstrated that the panicle branch length was significantly larger in transgenic lines with haplotype I than their corresponding null segregants ($P = 0.02$) (Fig. 3c). Transgenic lines with

haplotype I also had significantly longer branches than those carrying haplotype II ($P < 0.0001$). Finally, there was no significant difference between null segregants and transgenic lines with haplotype II ($P = 0.52$), or between null segregants originated from events with either haplotype ($P = 0.58$). These results indicate that panicle branch length was the major determinant of the observed seed number differences.

Other agronomic traits including plant height, flowering time, panicle exsertion, panicle length, and biomass, were also compared and these results suggest that *SbKS3* had pleiotropic effects on plant height since the same significant differences described for seed number, total seed weight per inflorescence, and panicle branch length were observed for plant height (Fig. 3d; Table 1).

Expression profile of *SbKS3* in sorghum

Two sets of experiments were set up to investigate the expression pattern of *SbKS3*. Two sorghum genotypes contrasting in seed number and *SbKS3* haplotypes were initially tested by RT-PCR on different tissues (root, leaf, stem, and meristem) and developmental times (three-leaf stage, the transition from vegetative to reproductive stage, and post-transition). The leaf was the only tissue with differential *SbKS3* expression between lines at the three-leaf stage, as shown in Fig. 4. When plants reached the transition from vegetative to reproductive growth stage (approximately seven-leaf), *SbKS3* was constitutively expressed in leaf, stem, and root tissue in both lines. After the transition, the plant enters a phase of reproductive growth critical for seed formation and the expression of *SbKS3* on leaf, stem, and root was dramatically reduced. Additionally, expression mostly occurred in the apical meristem of the high seed number accession, and reach a maximum in the immature panicle, at approximately 10-leaf stage. The

meristematic expression of the haplotype II allele was significantly lower with an alternative transcript generated only around the transition time.

An additional field validation study was conducted to determine the expression profile of *SbKS3* on an expanded set of ten sorghum lines. These results confirmed that *SbKS3* was only expressed in leaves of lines carrying the haplotype I at the three-leaf stage (Supplemental Fig. S3a). Consistent with the previous experiment, an alternative transcript was detected in meristematic tissue of low seed number lines during the transition time from vegetative to reproductive stage (Supplemental Fig. S3b). *SbKS3* was actively expressed in the meristem of both high and low seed number lines during the post-transition stage (Supplemental Fig. S3c).

Discussion

Comparative *in silico* analysis of *SbKS3* between and within species

Comparative genomic studies between sorghum, rice and maize, demonstrated that the *ent*-kaurene synthase family has a variable number of members (Supplemental Fig. S4). Orthologous of *SbKS3* have one copy in rice, but two copies in maize, which agrees with the whole genome duplication that occurred after sorghum and maize diverged from their common ancestor (Paterson et al. 2009). *SbKS3* belongs to a small gene family with three members in sorghum. Protein alignments between these three members confirmed that *SbKS3* is the shortest one with the N-terminal γ-domain missing in comparison with *SbKS1*, the longest *KS* protein (Supplemental Fig. S1). This result led to the hypothesis that *SbKS1* encodes the functional Ent-kaurene synthase, while *SbKS3* does not, and thus, *SbKS1* and *SbKS3* were tested for their *KS* activity through the same modular metabolic engineering system (Cyr et al. 2007). Contrary to the proposed hypothesis, both *SbKS1* and *SbKS3* had the catalytic capacity to synthesize Ent-kaurene from the precursor *Ent*-Copalyl diphosphate (Supplemental Fig. S5, Fig. 1).

SbKS1 and *SbKS3* form a tandem array in the long arm of sorghum chromosome 6, in agreement with a similar arrangement recently described in maize that includes *ZmKSL3, ZmTPS1,* and *ZmKSL5,* all with *KS* activity (Fu et al. 2016). *ZmKSL3*, the orthologous of *SbKS1*, serves as the main ent-kaurene synthase for the gibberellin metabolism in maize. *ZmTPS1* and *ZmKSL5* that have lost the N-terminal γ-domain, are the duplicated orthologous of *SbKS3* that have retained *KS* function. This tandem array of *ZmKSL3, ZmTPS1,* and *ZmKSL5* is syntenic to: a) sorghum chromosome 6, where *SbKS1* and *SbKS3* are located; and b) rice chromosome 4 that includes *Os06g06611700* and *Os06g06611800*. Therefore, members of this tandem duplication localized in syntenic regions may have similar functions in these different cereal species.

Comparative genomics of QTL related to seed number

There is a strong collinear relationship between the long arm of sorghum chromosome 6, rice chromosome 4 and maize chromosome 2, in the reverse orientation. Considering that *SbKS1*, *SbKS3* and their maize and rice orthologs are located in this syntenic region, the presence of QTL associated with seed number or grain yield in different species was investigated.

In rice, QTL for seed number and its components such as primary and secondary branch number, and panicle number per plant, were simultaneously identified within the syntenic region on the long arm of chromosome 4 (Deshmukh et al. 2010). This locus, called *qGN4-1*, was consistently discovered across three years with LOD scores of 13, 6.8, and 5.3, explaining 27, 16, and 12% of the phenotypic variation, respectively. This 0.78-Mb region on chromosome 4 is delimited by markers *nkssr04-02* and *nkssr04-19* and is located in the previously mentioned collinear region that contains the *SbKS3* rice orthologs.

QTL for seed number related traits were also detected in the same syntenic region in maize (Sabdin et al. 2008), wheat (Kuchel et al. 2007) and barley (Horsley et al. 2006). A locus

explaining 45% of the variation in kernel row number and ear number per plant was localized on bin 2.02 in maize (Sabdin et al. 2008). Similar discoveries were reported for wheat and barley, since the grain yield QTL *QGyld.agt-2D* was mapped to wheat chromosome 2D (Kuchel et al. 2007) and a spike density locus was localized on the syntenic chromosome 2 of barley (Horsley et al. 2006).

Syntenic studies and homologous cloning have been successfully applied for the identification of causal genes underlying traits of interest (Chardon et al. 2004; Kojima et al. 2002; Li et al. 2010; Yano et al. 2000). Therefore, the simultaneous identification of QTL related to seed number traits within this syntenic region across rice, maize, sorghum, wheat and barley is a strong evidence to support the hypothesis that *SbKS3* is involved in seed number determination.

SbKS3 has pleiotropic effects on seed number and plant height

Results from the *SbKS3* overexpression experiment in rice demonstrated that this gene controls both seed number and plant height (Table 1 and Fig. 4a and 4d). These pleiotropic effects have also been reported for other cloned genes affecting seed number (Xue et al. 2008; Wei et al. 2010). *Grain number, plant height and heading date 7* (*Ghd7*), encoding a CCT domain protein, has major effects on an array of traits in rice, including number of grains per panicle, plant height and heading date. Enhanced expression of *Ghd7* increased grains per panicle and plant height under long day conditions (Xue et al. 2008). Similar observations were reported by Wei et al. (2010) about *DAYS TO HEADING 8* (*DTH8*), another gene affecting seed number per panicle. Plant height was also significantly increased under long-day conditions when the *DHT8* functional allele was present, in comparison with the plant height response of

CSSL61 (a chromosome segment substitution line that carries the non-functional *DTH8* allele) (Wei et al. 2010).

The relationship between gibberellins and branch elongation

The role of *SbKS3* in determining seed number per inflorescence was confirmed by transformation experiments. Haplotype I, the favorable allele, increased seed number relative to the corresponding null segregants or transgenic lines carrying the haplotype II allele. Of all inflorescence architecture parameters investigated, panicle branch length was the only yield component altered by *SbKS3* overexpression and the morphological change behind the increased seed number. Considering that *SbKS3* was confirmed as a functional enzyme with the catalytic capacity to convert *Ent*-Copalyl diphosphate (substrate) into *Ent*-kaurene (product), we can hypothesize that *SbKS3* is affecting the final concentration of active GAs. GAs are an important group of plant hormones that regulate many aspects of plant development such as seed germination, stem growth, floral induction, and fruit growth (Kende and Zeevaart 1997; Olszewski et al. 2002; Yamaguchi 2008). Recently, Jiang et al (2014) discovered that GAs are also involved in the determination of inflorescence branching length. The "*sped1-D*" rice mutant was characterized as having shorter pedicels and secondary branches in comparison with the wild type phenotype. The expression of several *GID1L2* family members, important components of the GA signaling pathway, was downregulated in the *sped1-D* mutant, suggesting that the GA pathway is involved in branching elongation (Jiang et al. 2014). Therefore, our proposed hypothesis for further investigation is that *SbKS3* affects the accumulation of GAs, and thus panicle branch elongation, influencing the final number of seeds per inflorescence.

This hypothesis is also based on the observed *SbKS3* expression profile in sorghum that dramatically decreased in leaf, stem, and root after the transition from vegetative to reproductive

stage, and mainly occurred in the apical meristem at an important time for seed number determination. The maximum expression of *SbKS3* was observed when the meristem was between 4 cm and 12 cm for both high and low seed number lines (Fig. 3). According to Ikeda et al. (2004), when the panicle becomes longer than approx. 1.5 mm, each floret starts the floral organ differentiation and, floral organ primordia are formed in panicles 40 mm long. This process usually occurs in stage 7, in which all floral organs are set. Therefore, the number of primary and secondary branches will not change after this stage. Rachis and branches enter a phase of rapid elongation after the inflorescence becomes approx. 4 cm long, called stage 8 (inflorescence length 40-230 mm), in which the enlargement of organs and the ovule and pollen differentiation take place (Ikeda et al. 2004). The timing of maximum *SbKS3* expression relative to stages of inflorescence development explains why primary branch number was not the morphological change behind an increased seed number but was rather explained by changes in the length of panicle branches.

Acknowledgments

The authors appreciate Dr. Ruben Peters' lab members (Xuan Lü, Samuel Schulte) for their work on testing enzyme activity; Dr. Kan Wang and Bin Yang for rice transformation; Erica Unger-Wallace for meristem dissection; Jieyun Hu, Joshua Kemp, and Riqing Li for their support on both lab and greenhouse activities.

References

Ashikari M, Sakakibara H, Lin S, et al (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–5. doi: 10.1126/science.1113373

Bartrina I, Otto E, Strnad M, Werner T, Schmulling T (2011) Cytokinin Regulates the Activity of Reproductive Meristems, Flower Organ Size, Ovule Formation, and Thus Seed Yield in Arabidopsis thaliana. The Plant Cell 23: 69-80

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Boyles RE, Cooper EA, Myers MT, et al (2016) Genome-Wide Association Studies of Grain Yield Components in Diverse Sorghum Germplasm. Plant Genome 0:0. doi: 10.3835/plantgenome2015.09.0091

Chardon F, Virlon B, Moreau L, et al (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics 168:2169–2185. doi: 10.1534/genetics.104.032375

Cyr A, Wilderman PR, Determan M, Peters RJ (2007) A modular approach for facile biosynthesis of labdane-related diterpenes. J Am Chem Soc 129:6684–6685. doi: 10.1021/ja071158n

Deshmukh R, Singh A, Jain N, et al (2010) Identification of candidate genes for grain number in rice (Oryza sativa L.). Funct Integr Genomics 10:339–347. doi: 10.1007/s10142-010-0167-2

Fu J, Ren F, Lu X, et al (2016) A tandem array of ent-kaurene synthases in maize with roles in gibberellin and more specialized metabolism. Plant Physiol 170:742–751. doi: 10.1104/pp.114.253682

Gu B, Zhou TY, Luo JH, Liu H et al (2015) An-2 Encodes a Cytokinin Synthesis Enzyme that Regulates Awn Length and Grain Production in Rice. Molecular Plant 8, 1635–1650. doi.org/10.1016/j.molp.2015.08.001

Gerik, T., B. W. Bean, and R. Vanderlip. 2003. Sorghum Growth and Development. Agric. Communications, The Texas A&M Univ. Syst., College Station, TX.

Horsley RD, Schmierer D, Maier C, et al (2006) Identification of QTLs associated with fusarium head blight resistance in Barley Accession CIho 4196. Crop Sci 46:145–156. doi: 10.2135/cropsci2005.0247

Ikeda K, Sunohara H, Nagato Y (2004) Developmental Course of Inflorescence and Spikelet in Rice. Breeding Science 54:147-156

Jiang G, Xiang XH, Zhao JY, Yin DD, et al (2014) Regulation of Inflorescence Branch Development in Rice Through a Novel Pathway Involving the Pentatricopeptide Repeat Protein sped1-D. Genetics 197:1395–1407. doi: 10.1534/genetics.114.163931

Kende H, Zeevaart J (1997) The five “classical” plant hormones. Plant Cell 9: 1197–1210.

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to Xowering downstream of Hd1 under shortday conditions. Plant Cell Physiol 43(10):1096–1105

Kuchel H, Williams KJ, Langridge P, et al (2007) Genetic dissection of grain yield in bread wheat. I. QTL analysis. Theor Appl Genet 115:1029–1041. doi: 10.1007/s00122-007-0629-7

Li Q, Yang XH, Bai GH, Warburton WL, Mahuku G, Gore M, Dai JR, Li JS, Yan JB (2010) Cloning and characterization of a putative GS3 ortholog involved in maize kernel development. Theor Appl Genet 120:753–763 DOI 10.1007/s00122-009-1196-x

Milne I, Shaw P, Stephen G, et al (2010) Flapjack-graphical genotype visualization. Bioinformatics 26:3133–3134. doi: 10.1093/bioinformatics/btq580

Miura K, Ikeda M, Matsubara A, et al (2010) OsSPL14 promotes panicle branching and higher grain productivity in rice. Nat Genet 42:545–549. doi: 10.1038/ng.592

Morrone D, Lowry L, Determan MK, et al (2010) Increasing diterpene yield with a modular metabolic engineering system in E. coli: Comparison of MEV and MEP isoprenoid precursor pathway engineering. Appl Microbiol Biotechnol 85:1893–1906. doi: 10.1007/s00253-009-2219-x

Nagaraja Reddy R, Madhusudhana R, Murali Mohan S, et al (2013) Mapping QTL for grain yield and other agronomic traits in post-rainy sorghum [Sorghum bicolor (L.) Moench]. Theor Appl Genet 126:1921–1939. doi: 10.1007/s00122-013-2107-8

Olszewski N, Sun T P, Gubler F (2002) Gibberellin signaling: biosynthesis, catabolism, and response pathways. Plant Cell 14(Suppl.): S61–S80.

Ordonio RL, Ito Y, Hatakeyama A, et al (2015) Gibberellin deficiency pleiotropically induces culm bending in sorghum: an insight into sorghum semi-dwarf breeding. Sci Rep 4:5287. doi: 10.1038/srep05287

Paterson AH, Bowers JE, Bruggmann R, et al (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556. doi: 10.1038/nature07723

Peltonen-Sainio P, Kangas A, Salo Y, Jauhiainen L (2007) Grain number dominates grain weight in temperate cereal yield determination: Evidence based on 30 years of multi-location trials. F Crop Res 100:179–188. doi: 10.1016/j.fcr.2006.07.002

Peter H and Stephen T (2012) Gibberellin biosynthesis and its regulation. Biochem. J. 444, 11–25. doi:10.1042/BJ20120245

Rami J-F, Dufour P, Trouche G, et al (1998) Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (Sorghum bicolor L. Moench). TAG Theor Appl Genet 97:605–616. doi: 10.1007/s001220050936

Sabadin PK, Junior claudio lopes de souza, Souza anete pereira de, Garcia antonio augusto franco (2008) QTL mapping for yield components in a tropical maize population using microsatellite markers. Hereditas 145:194–203. doi: 10.1111/j.2008.0018-0661.02065.x

Saeed M, Francis CA, Clegg MD (1986) Yield Component Analysis in Grain Sorghum. Crop Sci 26:346–351. doi: 10.2135/cropsci1986.0011183X002600020028x

SAS Institute. 2009. User's guide: Statistics. SAS Inst., Cary, NC.

Sequencher® version 5.4.6 DNA sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA

Shin J-H, Blay S, Mcneney B, Graham J (2006) LDheatmap : An R Function for Graphical Display of Pairwise Linkage Disequiibria between Single Nucleotide Polymorphisms.

Si L, Chen J, Huang X, et al (2016) OsSPL13 controls grain size in cultivated rice. Nat Genet 48:447–456. doi: 10.1038/ng.3518

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539 doi:10.1038/msb.2011.75

Sreenivasulu N, Schnurbusch T (2012) A genetic playground for enhancing grain number in cereals. Trends Plant Sci 17:91–101. doi: 10.1016/j.tplants.2011.11.003

van Oosterom EJ, Hammer GL (2008) Determination of grain number in sorghum. F Crop Res 108:259–268. doi: 10.1016/j.fcr.2008.06.001

Wang X, Wang H, Liu S, et al (2016) Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. Nat Genet 48:1–12. doi: 10.1038/ng.3636

Wei X, Xu J, Guo H, et al (2010) DTH8 Suppresses Flowering in Rice, Influencing Plant Height and Yield Potential Simultaneously. Plant Physiol 153:1747–1758. doi: 10.1104/pp.110.156943

Wu Y, Wang Y, Mi XF, et al (2016) The QTL GNP1 Encodes GA20ox1, Which Increases Grain Number and Yield by Increasing Cytokinin Activity in Rice Panicle Meristems. PLoS Genet 12:1–21. doi: 10.1371/journal.pgen.1006386

Xue W, Xing Y, Weng X, et al (2008) Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat Genet 40:761–767. doi: 10.1038/ng.143

Yamaguchi S (2008) Gibberellin Metabolism and its Regulation. Annu Rev Plant Biol 59:225–251. doi: 10.1146/annurev.arplant.59.032607.092804

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis Xowering time gene CONSTANS. Plant Cell 12:2473–2484

Zawaski C, Kadmiel M, Pickens J, et al (2011) Repression of gibberellin biosynthesis or signaling produces striking alterations in poplar growth, morphology, and flowering. Planta 234:1285–1298. doi: 10.1007/s00425-011-1485-x

Zhang D, Li JP, Rosana O. Compton, JR, Goff VH, et al (2015) Comparative Genetics of Seed Size Traits in Divergent Cereal Lineages Represented by Sorghum (Panicoidae) and Rice (Oryzoidae). G3: GENES, GENOMES, GENETICS. 5:1117-1128. *doi.org/10.1534/g3.115.017590*

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler E (2010) Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42:355-360.

Zhao J, Mantilla Perez MB, Hu J, Salas Fernandez MG (2016) Genome-Wide Association Study for Nine Plant Architecture Traits in Sorghum. Plant Genome 9:0. doi: 10.3835/plantgenome2015.06.0044

Figures

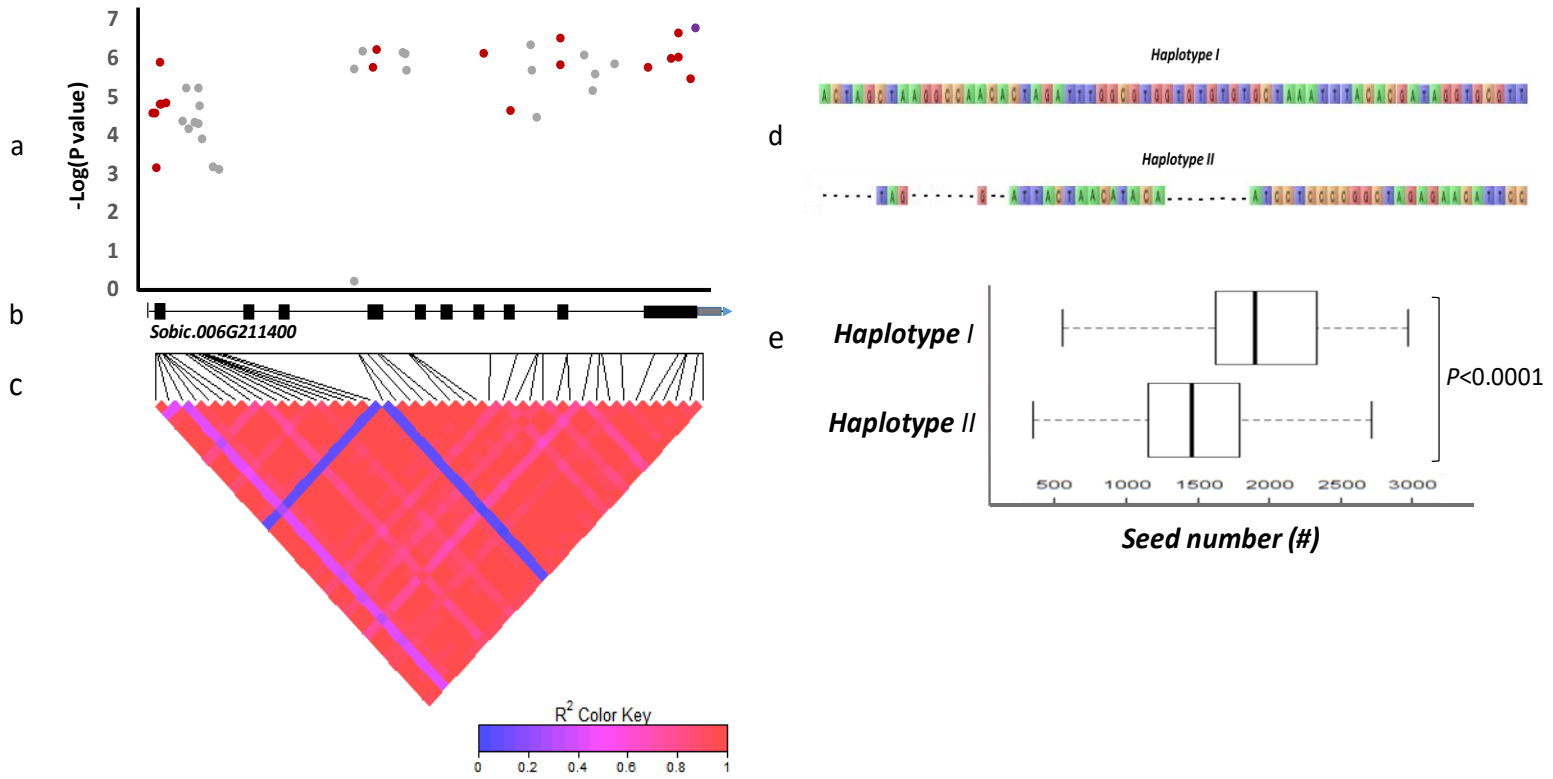
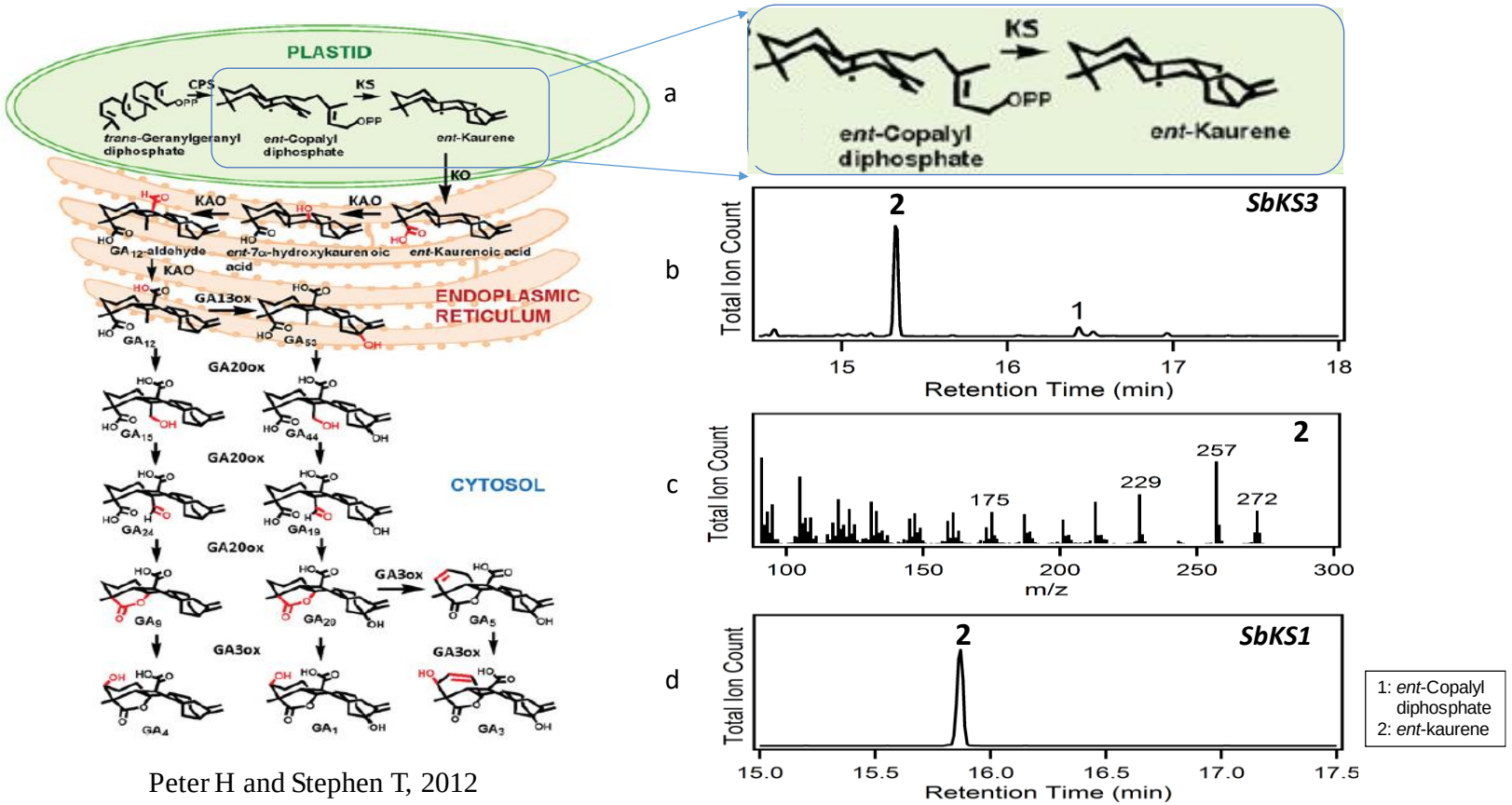


Figure 1. SbKS3-specific association analysis. (a) Association results of markers within SbKS3 exons (red dots), introns (gray dots), and 3' UTR (purple dot); (b) SbKS3 (Sobic.006g211400 /Sb06g028210) gene structure: exons (black box), introns (lines between black boxes), UTR (gray box); (c) LD plot of SbKS3 polymorphisms across the association panel; (d) SbKS3 alternative haplotypes; (e) Variation in seed number observed between the two alternative haplotypes.



Peter H and Stephen T, 2012

Figure 2. *SbKS1* and *SbKS3* biochemical recombination experiment. (a) *KS* functions in plastids, where it converts *ent*-Copalyl diphosphate into *ent*-kaurene (Peter H and Stephen T 2012); (b) GC-MS analysis of the enzymatic products of *SbKS3;* (c) Mass spectrum of *ent*-kaurene; (d) GC-MS analysis of the enzymatic products of *SbKS1*.

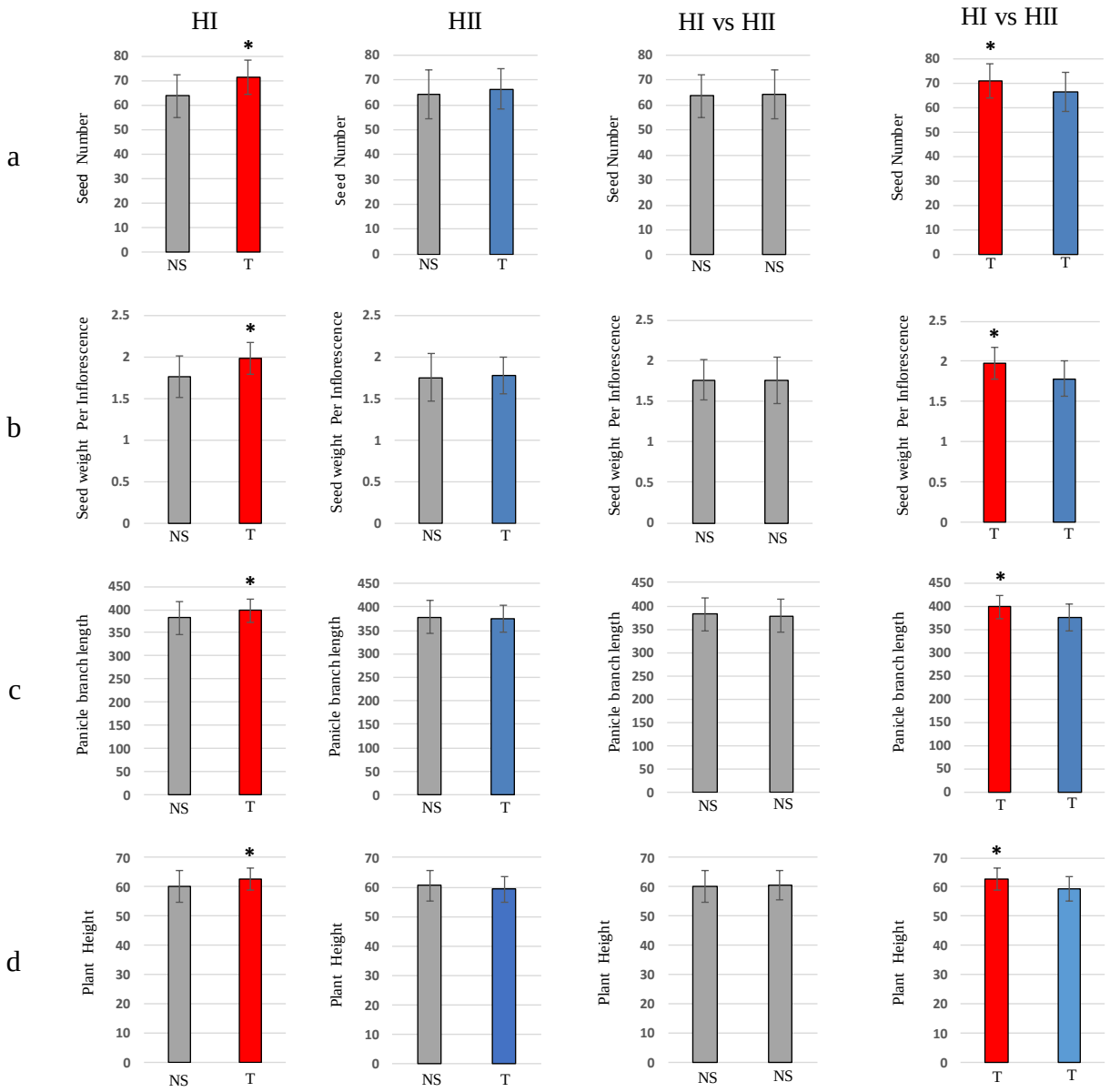


Figure 3. Phenotypic analysis of T_2 rice plants produced by overexpression of the two alternative *SbKS3* haplotypes. HI: haplotype I; HII: haplotype II. *represents the statistical significance at $P<0.05$. Red bar: average of 48 transgenic plants from three independent transformation events with haplotype I; Blue bar: average of 67 transgenic plants from three independent events with haplotype II; Gray bar: the average of 50 null segregant from three independent events of Haplotype I/ the average of 60 null segregants from three independent events of Haplotype II. Error bars represent standard deviation.

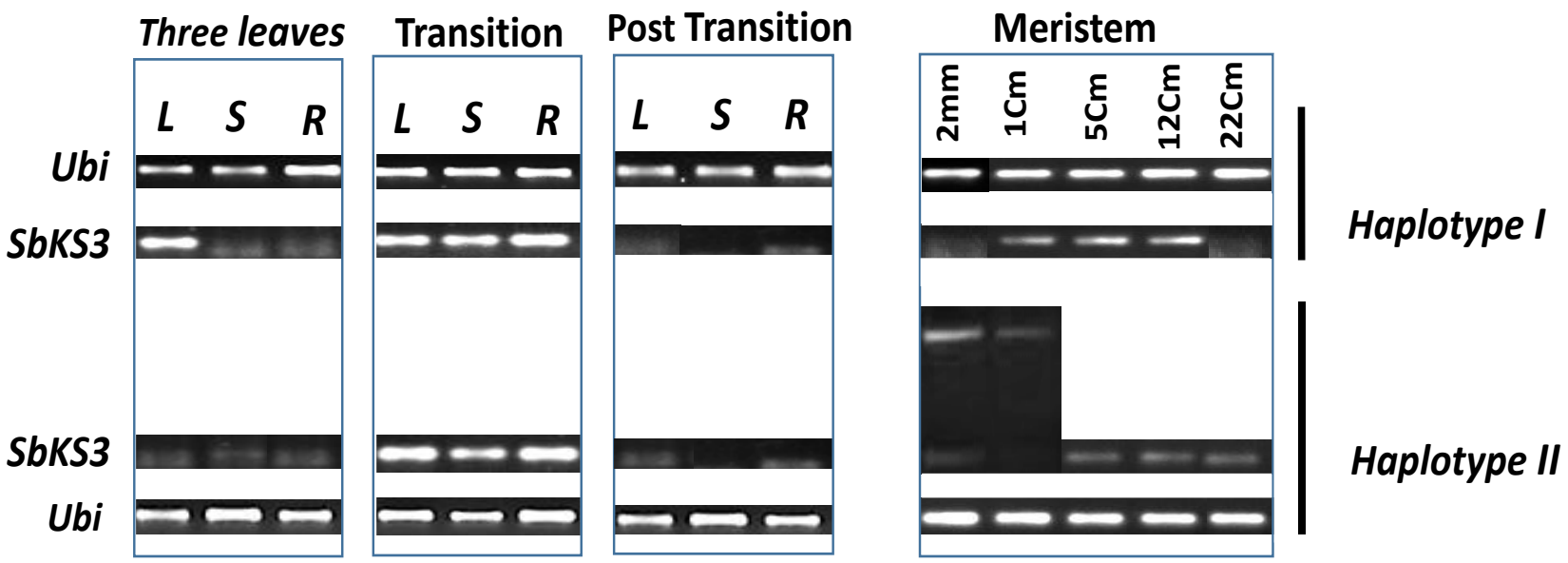


Figure 4. RT-PCR result of *SbKS3* on different tissues at different developmental stages. L: leaf tissue; S: stem tissue; R: root tissue at three-leaf stage, transition from vegetative to reproductive stage, post-transition (meristem approx. 4 cm long). RNA from the apical meristem was sampled at five time points (2 mm, 1 cm, 5 cm, 12 cm, 22 cm) from transition until the panicle was completely formed. Haplotype I corresponds to a high seed number line (PI552861), and haplotype II corresponds to a low seed number line (PI 533955). *Ubi*: Ubiquitin gene used as control.

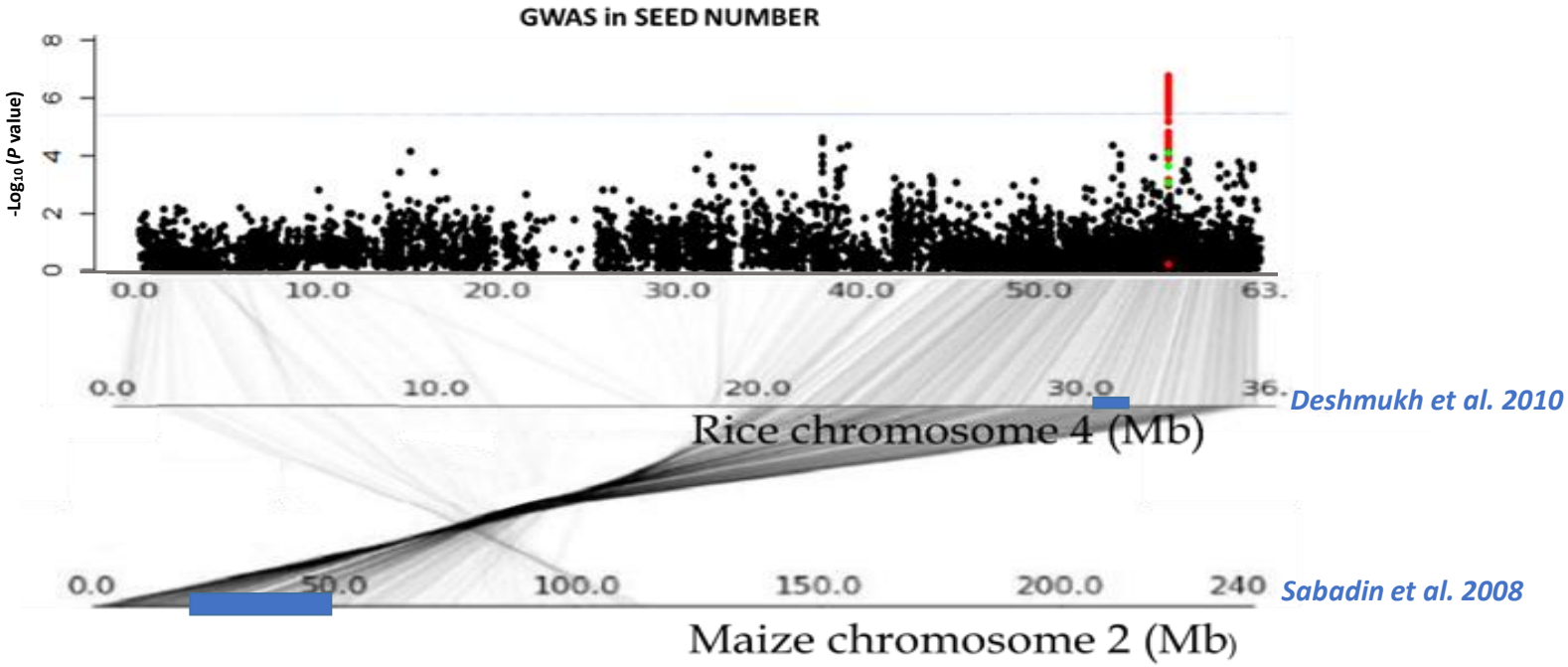


Figure 5. Syntenic relationship between sorghum chromosome 6, rice chromosome 4 and maize chromosome 2. Red dots represent SbKS3 polymorphisms; Green dots represent *SbKS1* polymorphisms. Blue boxes correspond to QTL on rice chromosome 4 (Deshmuk et al. 2010) and maize chromosome 2 (Sabadin et al. 2008).

Tables

Table 1. Statistical analysis of the *SbKS3* overexpression study in rice comparing the two alternative alleles and the corresponding null segregants for seed number, total seed weight per inflorescence, panicle branch length and plant height.

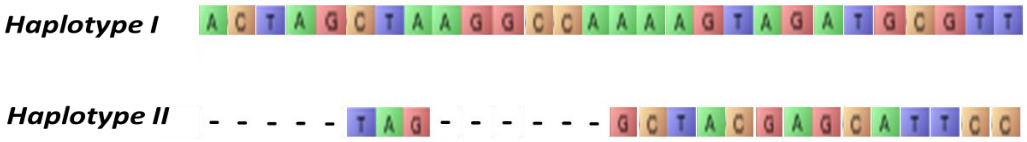
Trait	Haplotype	Comparison	*P* value	R^2
Seed number	H I	Null segregants vs Transgenic lines	<0.0001	0.19
Seed weight/inflorescence	H I	Null segregants vs Transgenic lines	<0.0001	0.22
Total branch length	H I	Null segregants vs Transgenic lines	0.02	0.06
Plant Height	H I	Null segregants vs Transgenic lines	0.009	0.07
Seed number	H II	Null segregants vs Transgenic lines	0.18	0.014
Seed weight/inflorescence	H II	Null segregants vs Transgenic lines	0.57	0.0025
Total branch length	H II	Null segregants vs Transgenic lines	0.52	0.003
Plant Height	H II	Null segregants vs Transgenic lines	0.14	0.02
Seed number	HI vs HII	Transgenic lines vs Transgenic lines	0.0014	0.089
Seed weight/inflorescence	HI vs HII	Transgenic lines vs Transgenic lines	<0.0001	0.19
Total branch length	HI vs HII	Transgenic lines vs Transgenic lines	<0.0001	0.15
Plant Height	HI vs HII	Transgenic lines vs Transgenic lines	<0.0001	0.16
Seed number	HI vs HII	Null segregants vs Null segregants	0.75	0.001
Seed weight/inflorescence	HI vs HII	Null segregants vs Null segregants	0.42	0.007
Total branch length	HI vs HII	Null segregants vs Null segregants	0.58	0.003
Plant Height	HI vs HII	Null segregants vs Null segregants	0.98	0.001

Supplemental data for Chapter 3

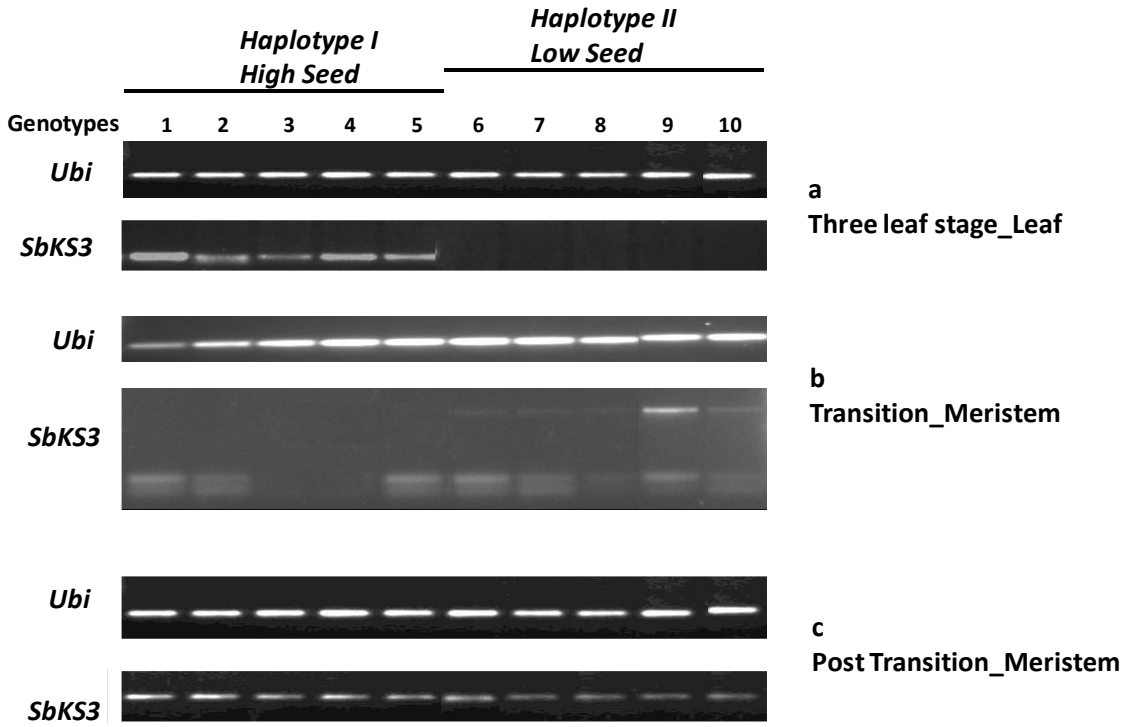
```
Sb05g022320      ------------------------------------------------------------
Sb06g028220      MPNVMAAAPAGCLPRGVIPPARSSRGRASLPGVARAYAERRLVAENTSLPNKHVQHKEEL
Sb06g028210      ------------------------------------------------------------


Sb05g022320      ------------------------MVPAQGDPHAPRFRGYVKWILENQHSDGSWGLGYLS
Sb06g028220      ETRIRNQLRRPQLPPSLYDTAWVSMVPMRGSHHTPCFPQCVEWILQNQQDDGSWGVSQSG
Sb06g028210      ------------------------------------------------------------


Sb05g022320      PRWLGKDCICSTLACILALKTWNIGDEHIRKGLHFIEKNSSYIMDEKSEAPVGFNIIFPN
Sb06g028220      -SEVNKDVLLSTLACVLALKRWNVGRENMWRGLHFIGRNFYVAMDEQVAAPVGFNITFPG
Sb06g028210      ------------------------------------------------------------


Sb05g022320      MIRLGIELGLEFPLKQSDFHQIFLLRE-MELQRITNSDSMALGRKAYMAYIAEGLADVLE
Sb06g028220      LLSLAIDMGLEIPIRQTDVCGILHLRE-MELKRQ--AVDSSYGRKTYMAYIAEGLGSMVD
Sb06g028210      ---------MAKPVKQLTAAASLSLMKAISSSSS--RLSSGGNAQRFGSSLPCGRG----
                                :  *::*      : * : :. .      . . . : : : :  * .

Sb05g022320      WDDVQKYQRKNGSLFNSPSATSAFA---IHTHNTNALKYLEFLGNKFVDSAPMAYPLSIQ
Sb06g028220      WDEVMKFQRKNGSLFSSPSTTAVAL---IHKYNDPA----------------LQYLNLLV
Sb06g028210      --RTMPTQRRSTSSSTRPAAPVNRVGPGRSKQHDK-----------------GASETTIM
                    .   **:. *  . *::          . :                          :

Sb05g022320      SQIFLVDILENMGISHRFSCEIKNILDMTYRFR----------------------------
Sb06g028220      SEFGSSDALERMGISQHFVSEIESILDMAYSCWLQKDEEIMMDKATCAMAFRLLRMNGYD
Sb06g028210      QQLQQVDVLENMGISRHFAGEIKRVLDRTYRCWLLRDEEIMLDAATCAMAFRILRMNGYD
                 .::   * **.****::*  **: :** :*

Sb05g022320      HATDVLSLFAEESRFHDSVEGHMNDTKALLELYKASLV-----------------------
Sb06g028220      VSSDVLSHVAGPSTFHDSLQGYLNDTKSLLELYKASKVSLSENDLVLDGIGFWSGNLLKD
Sb06g028210      VSSDELYHVAEASMFHNSLGGYLNDTRTMLELHKASTVSTSEDEYILDTIGSWSSTLLRE
                  ::* *  .*  * **:*: *::***:::***:*** *

Sb05g022320      -----------------EYALNLPFYSATLQPFKHKRNIECFGTEGIRIHKSAYLAC-DA
Sb06g028220      KLCSS-RVKKDLIFGEMEYAVKFPFYA-TLERLEHKRNIENFDAWGPLMLTTKS-SSFCI
Sb06g028210      QLGSGGALRRTPLFREVEHALDCPFYT-TLDRLDHRWNIENFNVTGHRMLETPYLSSRHT
                                 *:*:. ***: **: :.*: *** *.. *  :  :   :.

Sb05g022320      TENILALAIEDFHLSQSIYQQQLQYIERWVKEVRLDQLKFARDLPLSLFVFLATNVFPCE
Sb06g028220      DQEFVALAVEDFSFSQYVYQDELRHLDSWVKENKLDQLQFARQKLTYCYLSAAATIFSSE
Sb06g028210      SRDILTLAVRDFSSSQFKYQQELKHLESWVKECKLDQLPFARQKLAYFYLSAAGTMFPPE
                  .::::**:.**  **  **::*:::: **** :**** ***:     ::  * .:*  *

Sb05g022320      LYDASIAWTQKCILTTVVDDFFEGGGSTKELRNFVTLIEKWDMHAGIEFCSEDIEILFRA
Sb06g028220      LSDARISWAENGVLTTVVDDFFDVGGSKEELENLIALVEKWHAHHTVGFYSEQVKIVFSA
Sb06g028210      LSDARILWAKNGVLTTVVDDFFDVGGSKEELENLLLLVEMWDEHHKIEFYSEQVEIVFSS
                 * ** * *::: :*********: ***.:**.*:: *:* *. *  : * **:::*:* :

Sb05g022320      VYDTNNQIAAIGAKLQNRSVIDHIVEIWVKYVRTLMIEAEWTTKGHVPTMEEYMSVAETS
Sb06g028220      IYTTVNHLGVIASAAQGRDVTNHLVEIWLDLLRSMMVETEWQRSQYVPTVEEYMTNAVVS
Sb06g028210      IYNSVNQLGAKASLLQDRNVTKHLVQIWLDLLKSMMTEVEWRMSKYVPTEEEYMANASLT
                 :* : *::.. .:  *.*.* .*:*:**:. ::::* *.**  . :*** ****: *  :

Sb05g022320      SALGPVVVPSLYLVGPKLSDDMIRDPEYKNLLRYLGIGIRLINDIGTYEKEMSEGYVNSV
Sb06g028220      FALGPIVLPALYFVGQEVLEHAVKDEEYDELFRLMSTCGRLLNDSQSFEREGNQGKLNSV
Sb06g028210      FALGPIVLPTLYFLGPKIPKSAIKDPEYNELFRLMSTCGRLLNDVETFEREYNEGKLNSV
                  ****:*:*:**::* :: .  ::* **.:*:* :.   **:**  ::*:* .:* :***
```

```
Sb05g022320      LLRAFGDDAVMSPSSIEAAKREIHVLLANSQRELLKLVLNEGGPIPRPCKDIFWNTYKIG
Sb06g028220      SLLVRHSGGS---MSIEAAKKALQKSIDVSRRDLLRLVLRKESVVPRPCKELFWKMCKIL
Sb06g028210      SLLVLH-GGS---MSISDARRKLQKPIDTCRRDLLRLVLREEGVIPRPCKELFWKMCKVC
                  * .   ..     **. *:: ::  :  .:*:**:***.: . :*****::**:  *:

Sb05g022320      RQFYSEGDGFNMPQYLVAAVNAVIHEPLQQTPS-----------------------
Sb06g028220      HLFYFQTDGFSSPKEMVSAVNAVINEPLTVQNTTSSFLSSSSGK------------
Sb06g028210      YFFYSGTDGFSSPVEKAREVDAVINEPLKLQGSHASLRVVWEREEPSLNDVVHLIQ
                   **   ***. *   .  *:***:***    :
```

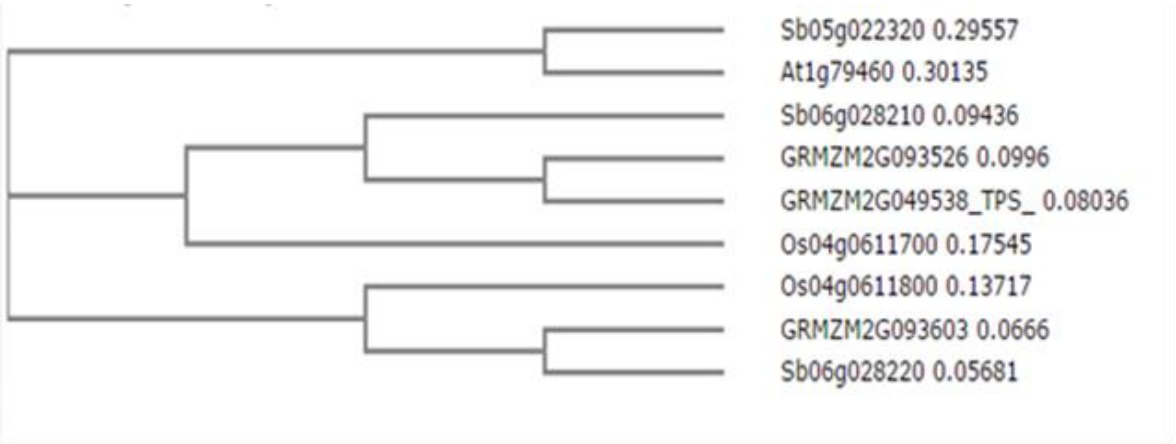
Supplemental Figure S1. Alignment of proteins from the three KS genes in sorghum. Sb06g028220: *SbKS1*; Sb05g022320: *SbKS2*; Sb06g028210: *SbKS3*.



Supplemental Figure S2. Polymorphisms of the two alternative haplotypes of *SbKS3* used for overexpression experiment in rice.



Supplemental Figure S3. RT-PCR results for ten sorghum lines with the two SbKS3 haplotypes. (a) Leaf tissue at the three-leaf stage. (b) Meristem tissue during the transition from vegetative to reproductive growth. (c) Meristem at the post-transition stage when the apical meristem was approx. 4 cm long.



Supplemental Figure S4. Phylogenetic tree of the KS gene family in maize, sorghum and rice.

Supplemental Table S1. Primer sequences

Gene	Forward primer	Reverse primer
SbKS3	TCCCGAAGTCTGCCATTAAAG	AGCTTGCCCTCATTGTACTC
SbKS1	GCATGCTGTCAAAGATGAAGAG	TGGCGGACAAGTAGAGAAAC
Ubiquitin	TTCTCAAAGAGCAGTGGAGC	TGTACATGTGGGCAATCTCAG

Supplemental Table S2. Single marker association analysis between seed number and polymorphisms in *SbKS3* using independent F2 populations and panicle length as a covariate.

Trait	**F2 Parents**	**Marker**	***P* value**	**R^2**	**Additive_*P***	**Dominant_*P***
Seed number	Btx2928 X 96Q63	S6_57009894	0.03	0.02	0.02	0.14
		S6_57009919	0.03	0.02	0.03	0.13
		S6_57009936	0.03	0.02	0.03	0.12
		S6_57009939	0.03	0.02	0.03	0.12
		S6_57009974	0.03	0.02	0.02	0.19
		S6_57010040	0.02	0.02	0.02	0.14
	Btx631 X SC648	S6_57009894	0.05	0.02	0.05	0.13
		S6_57009919	0.05	0.02	0.05	0.13
		S6_57009936	0.05	0.02	0.05	0.13
		S6_57009939	0.05	0.02	0.05	0.13
		S6_57009974	0.05	0.02	0.05	0.13
		S6_57010040	0.05	0.02	0.05	0.13

CHAPTER 4. CONCLUSIONS

This project has implemented a comprehensive approach to exploit natural genetic variation in sorghum for the discovery of genes of economic importance that would affect biomass and grain yield components. A GWAS was initially conducted as the first step to identify genomic regions and/or candidate genes controlling plant height, stem diameter, leaf angle, panicle length, panicle exsertion, flowering time, seed number, tiller number and internode number. Subsequently, the best candidate gene controlling seed number per inflorescence was further investigated by a series of complementary experiments to: i) confirm its effect on the target trait, ii) demonstrate its catalytic activity, iii) validate it using independent populations, iv) identify its orthologs in closely related species, v) characterize its expression profile, and vi) discover the mechanism used to exert an increase in seed number per panicle. These studies not only provide valuable biological information about the traits under investigation, but also one of the few examples in which a GWAS is successfully used as the first step in a gene discovery/cloning project.

In summary, our GWAS presented in chapter II has generated new knowledge about the genetic mechanisms underlying plant architecture parameters in sorghum, an important grain and bioenergy crop species. Nine traits, some of them highly correlated, were simultaneously investigated and our results compared with previously identified genomic regions using GWAS and linkage mapping. In summary, we have:

i. Discovered new genomic regions associated with all nine traits that could be further validated and narrowed down for their application in breeding programs.
ii. Confirmed previously identified QTL for some of the traits that provide independent validations to our results.

iii. Identified a few genomic regions that simultaneously control more than one trait, a phenomenon that could be due to pleiotropy or LD between different causal polymorphisms.

iv. Investigated in detail a leaf angle QTL that was reduced to a 1.67 Mb region with seven good candidate genes for future validation and cloning experiments.

Once our results are confirmed, those markers could be utilized in breeding programs to design the best sorghum ideotype for each environment and production system, and the knowledge could be additionally transferred to other important grass species such as rice, maize and wheat through comparative genomics.

The research described in chapter III was derived from the GWAS results presented in chapter II. *SbKS3,* first identified by GWAS as the candidate gene affecting seed number, has two alternative haplotypes that correspond to high and low seed number lines. Overall, we have:

i. Validated the association between allelic variants in SbKS3 and seed number per inflorescence.

ii. Confirmed the enzymatic activity of SbKS3 as an ent-kaurene synthase.

iii. Demonstrated that haplotype I, the favorable allele of SbKS3, increases seed number per panicle.

iv. Investigated the morphological changes associated with *SbKS3* effects on seed number using transgenic rice, and discovered a consistent increase in total branch length of the panicle.

v. Discovered the *SbKS3* expression pattern in multiple tissues and developmental stages that could be briefly described as differentially expressed in leaf tissue at three-leaf stage, then constitutively expressed in leaves at the vegetative-to-reproductive

transition stage, and later exclusively expressed on meristems after transition to reproductive growth.

The haplotype I of *SbKS3* could be utilized either by conventional breeding approaches or genetic manipulation strategies (genetic engineering or gene editing) to increase seed number not only in sorghum and rice but also in other important grass species such as maize and wheat, after testing in these species. According to yield data from the last century reported by the United States Department of Agriculture, National Agriculture Statistic Service (USDA-NASS), the average genetic grain has been approx. 1% per year for maize, rice and sorghum. Considering that *SbKS3* explained approx. 9% of the natural sorghum variation in seed number and that the favorable allele (Haplotype I) generated an increase in rice grain yield of similar magnitude, the impact of *SbKS3* on yield improvement could be significant. Elite sorghum germplasm with the unfavorable allele of *SbKS3* can be directly improved by substituting the unfavorable allele with haplotype I, increasing grain yield.

Although the function of *SbKS3* has been demonstrated herein, additional research is needed to identify the causal polymorphism/s to facilitate the implementation of gene editing techniques for crop improvement (Nekrasov et al. 2013). Additionally, our evidence suggests that the observed effect of haplotype I in sorghum could be the product of both a different protein from the one generated by haplotype II and the outcome of differential expression between these two alleles in different tissues and developmental stages. This dual mechanism (alternative protein sequences and expression profiles) deserves further investigation to facilitate the exploitation of this gene in its full potential. Our rice transformation study confirmed that amino acid differences between the alternative *SbKS3* alleles affect gene function, but the effect derived from expression differences is still unclear since *SbKS3* was constitutively expressed in

transgenic rice plants. This experimental difference between the transgenic study and the results obtained from natural variation in sorghum could also explain the observed effect of *SbKS3* on plant height of transgenic rice lines. This gene was not significantly associated with plant height in sorghum and minimizing undesirable pleiotropic effects is an important goal of any genetic engineering or gene editing efforts.

The alternative transcript discovered in sorghum for haplotype II during transition from vegetative growth to reproductive growth is another important result for further investigation. Alternative splicing refers to a process during gene expression that results in the synthesis of multiple proteins from a single gene (Black 2003). It plays important roles in regulating flowering (Lee et al. 2013), responses to stress (Liu et al. 2013), disease resistance (Torres et al. 2005) and regulation of hormonal responses (Wang et al. 2015). Understanding the underlying function of this alternative transcript from haplotype II may enrich our understanding of the *SbKS3* mechanism.

Complementary tests in other species will be important to confirm the function of *SbKS3* and its orthologs, and to exploit it in cereal species other than rice. E.g., the orthologue of *SbKS3* in other species could be knocked out first, and then a complementation study could be performed to confirm whether or not the mutant phenotype is recovered by overexpressing the cDNA of *SbKS3* (Yamaguhi et al. 1998).

Finally, protein-protein interactions or gene-gene interaction networks could be investigated to fully understand the biochemical and physiological mechanism of *SbKS3*. In rice, Wu and colleagues revealed that GA20ox1, a gibberellin biosynthetic enzyme, controls grain number and yield through a mechanism that involves a KNOX-mediated cytokinin and gibberellin crosstalk in the rice meristem (Wu et al. 2016). Therefore, it is possible that *SbKS3*

affects seed number per inflorescence and grain yield by crosstalk mechanisms with other hormonal groups including cytokinins.

References

Black, DL. (2003) Mechanisms of alternative pre-messenger RNA splicing". Annual Review of Biochemistry. 72 (1): 291-336. doi:10.1146/annurev.biochem.72.121801.161720.

Naqvi S, Farre G, Sanahuja G, Capell T, Zhu C, Christou P (2009) When more is better: multigene engineering in plants. Trends in Plant Science.15: 48-56.

Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotech. 31:691–693. doi:10.1038/nbt.2655.

Lee JH, Ryu HS, Chung KS, Posé D, Kim S, Schmid M, Ahn JH (2013) Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. Science. 342:628–632. doi: 10.1126/science.1241097.

Liu J, Sun N, Liu M, Liu J, Du B, Wang X, Qi X (2013) An autoregulatory loop controlling *Arabidopsis*HsfA2 expression: Role of heat shock-induced alternative splicing. Plant Physiol. 162:512–521. doi: 10.1104/pp.112.205864.

Tanaka, Y, Brugliera F, Chandler S (2009) Recent Progress of flower colour modification by biotechnology. Int. J. Mol. Sci. 10: 5350 – 5369.

Torres MA, Jones JD, Dangl JL (2005) Pathogen-induced, NADPH oxidase-derived reactive oxygen intermediates suppress spread of cell death in *Arabidopsis thaliana*. Nat. Genet. 37:1130–1134. doi: 10.1038/ng1639.

Wang Z, Ji H, Yuan B, Wang S, Su C, Yao B, Zhao H, Li X (2015) ABA signaling is fine-tuned by antagonistic HAB1 variants. Nat. Commun. 6:1–12. doi: 10.1038/nature14444.

Wu Y, Wang Y, Mi XF, et al (2016) The QTL GNP1 Encodes GA20ox1, Which Increases Grain Number and Yield by Increasing Cytokinin Activity in Rice Panicle Meristems. PLoS Genet 12:1–21. doi: 10.1371/journal.pgen.1006386.

Yamaguchi S, Sun T, Hiroshi K, Kamiya Y (1998) The GA2 Locus of Arabidopsis thaliana Encodes ent-Kaurene Synthase of Gibberellin Biosynthesis. Plant Physiol. 116: 1271–1278.

Ye XD, Babili A, Kloti S, Zhang A, Lucca J, Beyer P, Potrykus I (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science. 287: 303–305.